EXHIBIT B

Trials@uspto.gov
571-272-7822

Paper 57
Date: May 11, 2022

UNITED STATES PATENT AND TRADEMARK OFFICE

———————

BEFORE THE PATENT TRIAL AND APPEAL BOARD

———————

GOOGLE LLC,
Petitioner,

v.

SINGULAR COMPUTING LLC,
Patent Owner.

———————

IPR2021-00179
Patent 8,407,273 B2

———————

Before JUSTIN T. ARBES, STACEY G. WHITE, and JASON M. REPKO,
Administrative Patent Judges.

PER CURIAM.

JUDGMENT
Final Written Decision
Determining Some Challenged Claims Unpatentable
35 U.S.C. § 318(a)
Dismissing Patent Owner's Motion to Exclude
37 C.F.R. § 42.64
Granting Patent Owner's and Petitioner's Motions to Seal
37 C.F.R. §§ 42.14, 42.54

IPR2021-00179
Patent 8,407,273 B2

I.   INTRODUCTION

A.   Background and Summary

Petitioner Google LLC filed a Petition (Paper 2, "Pet.") requesting *inter partes* review of claims 1–26, 28, 32–61, 63, and 67–70 of U.S. Patent No. 8,407,273 B2 (Ex. 1001, "the '273 patent") pursuant to 35 U.S.C. § 311(a). On May 14, 2021, we instituted an *inter partes* review as to all challenged claims on all grounds of unpatentability asserted in the Petition. Paper 16 ("Decision on Institution" or "Dec. on Inst."). Patent Owner Singular Computing LLC subsequently filed a Patent Owner Response (Paper 28, "PO Resp."), Petitioner filed a Reply (Paper 34, "Reply"), and Patent Owner filed a Sur-Reply (Paper 39, "Sur-Reply"). Patent Owner filed a Motion to Exclude (Paper 46, "Mot.") certain evidence submitted by Petitioner, to which Petitioner filed an Opposition (Paper 47) and Patent Owner filed a Reply (Paper 49). An oral hearing was held on February 11, 2022, and transcripts of the hearing are included in the record (Paper 55, "Conf. Tr."; Paper 56, "Public Tr."). The parties also filed Motions to Seal (Papers 52 and 54) portions of their demonstrative exhibits containing confidential information.

We have jurisdiction under 35 U.S.C. § 6. This Final Written Decision is issued pursuant to 35 U.S.C. § 318(a). For the reasons that follow, we determine that Petitioner has shown by a preponderance of the evidence that claims 1, 2, 21–24, 26, 28, 32, and 33 are unpatentable, but has not shown by a preponderance of the evidence that claims 3–20, 25, 34–61, 63, and 67–70 are unpatentable.

IPR2021-00179
Patent 8,407,273 B2

B.   Related Matters

The parties indicate that the '273 patent is the subject of *Singular Computing LLC v. Google LLC*, Case No. 1:19-cv-12551-FDS (D. Mass.) ("the district court case"). *See* Pet. xi; Paper 8, 1. Petitioner filed another petition challenging claims 1–70 of the '273 patent in IPR2021-00178, which was denied, and filed four other petitions challenging claims of two related patents also asserted in the district court case in IPR2021-00154 (denied), IPR2021-00155 (instituted), IPR2021-00164 (denied), and IPR2021-00165 (instituted).

C. The '273 Patent

The '273 patent, entitled "Processing with Compact Arithmetic Processing Element," issued on March 26, 2013. Ex. 1001, codes (45), (54). The '273 patent describes "computer processors or other devices which use low precision high dynamic range (LPHDR) processing elements to perform computations (such as arithmetic operations)." *Id.* at col. 5, l. 65–col. 6, l. 2. According to the '273 patent, conventional central processing unit (CPU) chips make inefficient use of transistors as a tradeoff for delivering the high precision required by many applications. *Id.* at col. 3, ll. 7–22. For example, conventional CPU chips "perform[] exact arithmetic with integers typically 32 or 64 bits long and perform[] rather accurate and widely standardized arithmetic with 32 and 64 bit floating point numbers," but require "on the order of a million transistors to implement the arithmetic operations." *Id.* at col. 3, ll. 15–22. According to the '273 patent, "many economically important applications . . . are not especially sensitive to precision and . . . would greatly benefit, in the form of application performance per transistor, from the ability to draw upon a far greater

3

IPR2021-00179
Patent 8,407,273 B2

fraction of the computing power inherent in those million transistors." *Id.* at col. 3, ll. 23–28. But "[c]urrent architectures for general purpose computing fail to deliver this power." *Id.* at col. 3, ll. 28–29.

The '273 patent is, therefore, "directed to a processor or other device, such as a programmable and/or massively parallel processor or other device, which includes processing elements designed to perform arithmetic operations . . . on numerical values of low precision but high dynamic range ('LPHDR arithmetic')." *Id.* at col. 2, ll. 11–18. According to the '273 patent, "'low precision' processing elements perform arithmetic operations which produce results that frequently differ from exact results by at least 0.1%." *Id.* at col. 2, ll. 28–31. In addition, "high dynamic range" processing elements "are capable of operating on inputs and/or producing outputs spanning a range at least as large as from one millionth to one million." *Id.* at col. 2, ll. 35–39.

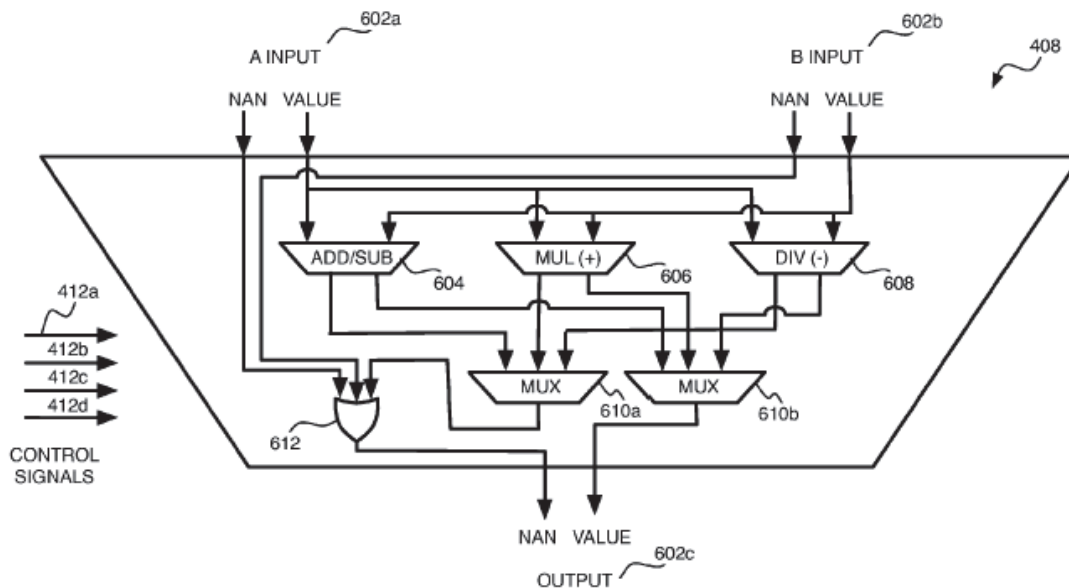Figure 6 of the '273 patent is reproduced below.

FIG. 6

4

IPR2021-00179
Patent 8,407,273 B2

Figure 6 depicts "an example design for an LPHDR arithmetic unit according to one embodiment of [the '273 patent]." *Id.* at col. 2, ll. 56–57. As shown in Figure 6, LPHDR arithmetic unit 408 receives two inputs: A input (602a) and B input (602b), and produces an output 602c. *Id.* at col. 12, ll. 52–53. The LPHDR arithmetic unit "is controlled by control signals 412a-d, coming from the CU 106, that determine which available arithmetic operation will be performed on the inputs 602a-b." *Id.* at col. 12, ll. 59–62. According to the '273 patent, Figure 6 illustrates an embodiment where "all the available arithmetic operations are performed in parallel on the inputs 602a-b by adder/subtractor 604, multiplier 606, and divider 608." *Id.* at col. 12, ll. 62–65. Finally, multiplexers (MUXes) 610a and 610b choose and send the desired result from among the outputs of the adder/subtractor, multiplier, and divider to output 602c. *Id.* at col. 13, ll. 1–8.

The '273 patent provides that "[t]he computing architecture literature discusses many variations which may be incorporated into the embodiment illustrated in FIG. 6." *Id.* at col. 13, ll. 8–10. According to the '273 patent, the "computational tasks" that the LPHDR arithmetic units can perform "enable a variety of practical applications." *Id.* at col. 17, ll. 18–21. The '273 patent provides, as examples, applications including "finding nearest neighbors," *id.* at col. 17, l. 29–col. 21, l. 32, "distance weighted scoring," *id.* at col. 21, l. 34–col. 22, l. 24, and "removing motion blur in images," *id.* at col. 22, l. 26–col. 23, l. 30.

IPR2021-00179
Patent 8,407,273 B2

D. Illustrative Claims

Challenged claims 1, 33, 36, and 68 of the '273 patent are independent. Claims 2–26, 28, and 32 depend from claim 1, claims 34 and 35 depend from claim 33, claims 37–61, 63, and 67 depend from claim 36, and claims 69 and 70 depend from claim 68. Claims 1 and 8 recite:

> 1. A device:

> comprising at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,

> wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/65,000 through 65,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input.

> 8. The device of claim 1, wherein the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide.

E. Evidence

The pending grounds of unpatentability in the instant *inter partes* review are based on the following prior art:

> U.S. Patent No. 5,689,677, issued Nov. 18, 1997 (Ex. 1009, "MacMillan");

> U.S. Patent Application Publication No. 2007/0203967 A1, published Aug. 30, 2007 (Ex. 1007, "Dockser"); and

IPR2021-00179
Patent 8,407,273 B2

>Jonathan Ying Fai Tong, David Nagle, & Rob A. Rutenbar, "Reducing Power by Optimizing the Necessary Precision/Range of Floating-Point Arithmetic," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 8, no. 3 (June 2000) (Ex. 1008, "Tong").

Petitioner filed a declaration from Richard Goodin, P.E. (Ex. 1003) with its Petition and a reply declaration from Mr. Goodin (Ex. 1071) with its Reply. Patent Owner filed declarations from Sunil P. Khatri, Ph.D. (Ex. 2051) and Joseph Bates, Ph.D. (the named inventor of the '273 patent) (Ex. 2052) with its Response. Also submitted as evidence are transcripts of the depositions of Mr. Goodin (Ex. 2043) and Dr. Khatri (Ex. 1072).

F. Asserted Grounds

This *inter partes* review involves the following grounds of unpatentability:

Claims Challenged	35 U.S.C. §	Reference(s)/Basis
1, 2, 21–24, 26, 28	103(a)[1]	Dockser
1, 2, 21–24, 26, 28, 32, 33	103(a)	Dockser, Tong
1–26, 28, 36–61, 63	103(a)	Dockser, MacMillan
1–26, 28, 32–61, 63, 67–70	103(a)	Dockser, Tong, MacMillan

---

[1] The Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011) ("AIA"), amended 35 U.S.C. § 103. Here, Petitioner's challenges are based on the pre-AIA version of 35 U.S.C. § 103.

IPR2021-00179
Patent 8,407,273 B2

II. ANALYSIS

A. *Level of Ordinary Skill in the Art*

In determining the level of ordinary skill in the art for a challenged patent, we look to "1) the types of problems encountered in the art; 2) the prior art solutions to those problems; 3) the rapidity with which innovations are made; 4) the sophistication of the technology; and 5) the educational level of active workers in the field." *Ruiz v. A.B. Chance Co.*, 234 F.3d 654, 666–667 (Fed. Cir. 2000). "Not all such factors may be present in every case, and one or more of them may predominate." *Id.*

Petitioner asserts that at the time of the earliest possible effective filing date of the '273 patent (June 19, 2009), a person of ordinary skill in the art would have had "at least a bachelor's degree in Electrical Engineering, Computer Engineering, Applied Mathematics, or the equivalent, and at least two years of academic or industry experience in computer architecture." Pet. 7–8 (citing Ex. 1003 ¶¶ 43–45). Patent Owner's proposed definition of the level of ordinary skill in the art is the same other than it removes the phrases "or the equivalent" and "at least." PO Resp. 14 (citing Ex. 2001 ¶¶ 36–37). There is little difference between the parties' proposed definitions, and the parties and their experts do not explain the basis for their proposals. Arguably, however, the term "at least" creates unnecessary ambiguity. Based on the full record developed during trial, including our review of the '273 patent and the types of problems and solutions described in the '273 patent and cited prior art, we determine that a person of ordinary skill in the art would have had a bachelor's degree in Electrical Engineering, Computer Engineering, Applied Mathematics, or the equivalent, and two years of academic or industry experience in computer architecture. *See, e.g.*, Ex. 1001, col. 1, l. 26–col. 2, l. 7 (describing in the

IPR2021-00179
Patent 8,407,273 B2

"Background" section of the '273 patent various conventional methods of computation and their alleged deficiencies). We apply that definition for purposes of this Decision.

B.   Claim Interpretation

We interpret the claims of the challenged patent

> using the same claim construction standard that would be used to construe the [claims] in a civil action under 35 U.S.C. 282(b), including construing the [claims] in accordance with the ordinary and customary meaning of such [claims] as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

37 C.F.R. § 42.100(b) (2020). "In determining the meaning of [a] disputed claim limitation, we look principally to the intrinsic evidence of record, examining the claim language itself, the written description, and the prosecution history, if in evidence." *DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.*, 469 F.3d 1005, 1014 (Fed. Cir. 2006). Claim terms are given their plain and ordinary meaning as would be understood by a person of ordinary skill in the art at the time of the invention and in the context of the entire patent disclosure. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313 (Fed. Cir. 2005) (en banc). "There are only two exceptions to this general rule: 1) when a patentee sets out a definition and acts as his own lexicographer, or 2) when the patentee disavows the full scope of a claim term either in the specification or during prosecution." *Thorner v. Sony Comput. Entm't Am. LLC*, 669 F.3d 1362, 1365 (Fed. Cir. 2012). We conclude that only one claim term requires interpretation to decide the issues presented during trial. *See Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017) ("Because we need only

IPR2021-00179
Patent 8,407,273 B2

construe terms 'that are in controversy, and only to the extent necessary to resolve the controversy,' we need not construe [a particular claim limitation] where the construction is not 'material to the . . . dispute.'" (citation omitted)).

Claim 1 recites "at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value." Independent claims 33, 36, and 68 recite the same limitation.

At issue here is the meaning of "low precision high dynamic range (LPHDR) execution unit," as recited in claims 1, 33, 36, and 68. Specifically, we need to resolve the question of whether this execution unit must be adapted to execute arithmetic operations *only* at low precision. For the reasons that follow, we determine that it does not.

Petitioner argues that, in the prior art used in all challenges, Dockser's floating-point processor (FPP), including the floating-point operator (FPO) inside the FPP, is low precision "because 'the precision' of operations in the FPP is 'reduced,' . . . and because it operates with the minimum imprecision" recited in the claim. Pet. 13–14 (citing Ex. 1003 ¶¶ 217–218; Ex. 1001, col. 26, l. 50–col. 27, l. 62; Ex. 1035, 8). Petitioner asserts that, in the claim, the minimum imprecision is found in the limitation that it calls *1B2*:

> for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least

10

IPR2021-00179
Patent 8,407,273 B2

> Y=0.05% from the result of an exact mathematical calculation of
> the first operation on the numerical values of that same input.

Ex. 1001, col. 30, ll. 6–16. According to Petitioner, the '273 patent

characterizes degrees of precision described in language mirroring *1B2* as

"low precision." Pet. 13–14. Petitioner argues that the execution units in

the prior art fall within the scope of the recited unit if they perform reduced

precision operations and at least one, but not necessarily all, of those

operations meets the other requirements of claim 1. *See, e.g., id.* at 30

(explaining that "numerous obvious implementations of Dockser's FPP drop

sufficient bits to yield an execution unit meeting the claimed minimum

imprecision"); Reply 2–3 (discussing the "low precision" requirement).

Patent Owner argues that the LPHDR execution unit should be

construed as "an execution unit that executes arithmetic operations *only* at

low precision and with high dynamic range, wherein 'high dynamic range'

and 'low precision' are defined according to the numerical requirements

below." PO Resp. 15 (emphasis added). We emphasize "only" because

Patent Owner argues that the recited LPHDR execution unit "necessarily

excludes full-precision or mixed full and low precision units." *Id.* at 16

(citing Ex. 2051 ¶ 60).

In support of its proposed construction, Patent Owner asserts that

construing the LPHDR execution unit as encompassing full or mixed

precision units would "impermissibly read out 'low precision high dynamic

range' and render the limitation meaningless." *Id.* (citing *Network-1 Techs.,

Inc. v. Hewlett-Packard Co.*, 981 F.3d 1015, 1022–23 (Fed. Cir. 2020)); *see

also id.* at 17 (arguing that Petitioner's construction reads "'low precision'

out of the claims entirely").

11

IPR2021-00179
Patent 8,407,273 B2

Based on the totality of the record, we agree with Petitioner that the recited LPHDR execution unit does not exclude mixed full and low precision execution units (i.e., those that are capable of arithmetic operations at both levels of precision). *See, e.g.*, Pet. 13–14, 30. We reject Patent Owner's argument that interpreting the claim in this way renders the term "low precision" "meaningless." *See* PO Resp. 15–17. Rather, the term "low precision" imposes the requirement that the unit must *at least* perform arithmetic operations at low precision, which would exclude units that operate only at full precision, for example. Thus, we disagree with Patent Owner and assign little weight to Dr. Khatri's testimony on this issue. *See id.*; Ex. 2051 ¶¶ 60–61.

As stated in our Decision on Institution, the recitation of "a first operation" in claim 1 requires only one or more first operations that meet the low precision criteria specified in the claim. Dec. on Inst. 19. The claims do not require that "every" operation must be low precision or exclude the LPHDR execution unit from having other capabilities for other operations. *Id.* Indeed, Patent Owner agrees that "a claim's recitation of an 'execution unit adapted to execute a first operation' would normally not preclude that execution unit from performing other types of operations." PO Resp. 17–18. Patent Owner, though, argues that this "rule does not apply when the specification teaches otherwise, as is the situation here." *Id.* We disagree.

Patent Owner relies on exemplary non-limiting embodiments from the '273 patent. *See id.* at 18. Specifically, Patent Owner argues that "the claimed LPHDR execution units are smaller and have fewer transistors" than prior art full precision execution units. *Id.* (citing Ex. 2051 ¶ 62). But the cited passage pertains only to exemplary embodiments and says that the

12

IPR2021-00179
Patent 8,407,273 B2

"amount of resources" is small, with transistors being only one example of such resources:

> For example, embodiments of the present invention may be implemented as any kind of machine which uses LPHDR arithmetic processing elements to provide computing using a *small amount of resources (e.g., transistors or volume)* compared with traditional architectures.

Ex. 1001, col. 8, ll. 3–7 (emphasis added). In fact, some embodiments do not even require transistors: "[e]xamples of such technologies include . . . other technologies whether based on transistors or not." *Id.* at col. 26, ll. 24–31, *cited in* Reply 6.

The reference to the "small amount of resources" in the cited passage (Ex. 1001, col. 8, ll. 3–7) is consistent with Petitioner's position that the LPHDR execution unit need only perform low precision operations some of the time because the record shows that doing so would save power and other computational resources. For example, Dr. Khatri testified that "exact computing . . . takes up a lot of resources, takes up a lot of power, takes up a lot of area, and takes up a lot of delay." Ex. 1072, 47:2–5. Mr. Goodin explained that reducing precision may achieve power savings. *See, e.g.,* Ex. 2043, 34:21–35:4; *see also id.* at 30:22–31:9 (discussing how selectable precision may achieve power savings). Mr. Goodin cites specific examples of how this is achieved. *See, e.g.,* Ex. 1003 ¶¶ 31–33 (citing Ex. 1008, 273, 277–279), 35 (citing Ex. 1007 ¶¶ 3–7). Thus, we disagree with Patent Owner's argument about reducing transistors and assign little weight to Dr. Khatri's testimony on this issue. *See* PO Resp. 18; Ex. 2051 ¶ 62.

Also, we agree with Petitioner that the degree of low precision is described in limitation *1B2* in claim 1 and similar limitations in the other claims. Pet. 13–14. The language used in *1B2* is consistent with how the

IPR2021-00179
Patent 8,407,273 B2

'273 patent describes low precision. Ex. 1001, col. 26, l. 50–col. 27, l. 62, *cited in* Pet. 13–14. For example, the patent states, "[t]he degree of precision of a 'low precision, high dynamic range' arithmetic element may vary from implementation to implementation." *Id.* at col. 26, ll. 50–52. The patent then lists several examples. *Id.* at col. 26, l. 52–col. 27, l. 62. In one example, "a LPHDR arithmetic element produces results which are sometimes (or all of the time) no closer than 0.05% to the correct result." *Id.* at col. 26, ll. 55–60. Similarly, claim 1 recites (and claims 33, 36, and 68 recite similar limitations) that

> the statistical mean . . . of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input.

Patent Owner agrees that the claims provide the parameters of the low precision and high dynamic range in the limitations "wherein the dynamic range . . ." and "for at least X=5% . . . ." PO Resp. 16. Indeed, Patent Owner in the district court case argued that the subsequent claim language specifying percentages "*defines* the term 'low precision'" and "specifies the degree of precision that an LPHDR execution unit must have in order to satisfy the 'low precision' . . . limitation of the claim." Ex. 1065, 9 (emphasis added); *see also* Ex. 1066, 12 (arguing that "'low precision high dynamic range' is defined in [the] claim itself"); Reply 2. Patent Owner, however, argues that those limitations "do not allow for the execution unit to have full-precision (or [low] dynamic range) capability." PO Resp. 16.

But the claims do not recite that every operation must meet these criteria. Rather, the claims only recite that the criteria apply to "a first operation." This reading is consistent with the examples of LPHDR

14

IPR2021-00179
Patent 8,407,273 B2

elements from the written description. *See* Ex. 1001, col. 26, l. 50–col. 27, l. 62, *cited in* Pet. 13–14. In those examples, the LPHDR arithmetic elements may produce results that are only "sometimes" no closer than the percentage to the correct result. *See id.* Thus, we agree with Patent Owner's argument that the claims provide the parameters for the term "low precision," but we disagree that the parameters must apply to every operation. *See* PO Resp. 16. For the same reasons, we assign little weight to Dr. Khatri's testimony on this issue. *See* Ex. 2051 ¶¶ 60–64.

Patent Owner also argues that "small size and low transistor count . . . is what allows a much larger number of LPHDR execution units to operate in parallel on a single chip." PO Resp. 19. But the requirement of a "much larger number" is not in the independent claims. Nor do the claims (or Patent Owner's proposed construction) include any limitations on size of the LPHDR execution unit or its transistor count. And "not every benefit flowing from an invention is a claim limitation." *i4i Ltd. P'ship v. Microsoft Corp.*, 598 F.3d 831, 843 (Fed. Cir. 2010). Even so, the Specification states that some embodiments are not based on transistors. *See, e.g.*, Ex. 1001, col. 26, ll. 24–31. So we find no support in the written description for Patent Owner's argument about small size and low transistor count. *See* PO Resp. 19. For the same reasons, we assign little weight to Dr. Khatri's testimony on this issue. *See* Ex. 2051 ¶¶ 63–64.

We also agree with Petitioner that the written description does not support Patent Owner's assertion that the '273 patent contrasts LPHDR execution units with other units that operate in both high and low precision. *See* PO Resp. 19, 23 (citing Ex. 1001, col. 5, ll. 31–40); Reply 7 (discussing Patent Owner's argument). The relevant part of the '273 patent describes a graphics processor that supports 16-bit and 32-bit floating point formats:

IPR2021-00179
Patent 8,407,273 B2

> When a graphics processor includes support for 16 bit floating point, that support is alongside support for 32 bit floating point, and increasingly, 64 bit floating point. That is, the 16 bit floating point format is supported for those applications that want it, but the higher precision formats also are supported because they are believed to be needed for traditional graphics applications and also for so called "general purpose" GPU applications. Thus, existing GPUs devote substantial resources to 32 (and increasingly 64) bit arithmetic and are wasteful of transistors in the sense discussed above.

Ex. 1001, col. 5, ll. 31–40. This paragraph does not define the term "low precision." *Id.* Nor does the patent explain how the 16-bit units meet the claimed requirements for low precision units that are recited in the claim. *See id.* Also, there is no discussion of these units in the part of the written description that lists examples of low precision units. *See id.* at col. 26, l. 50–col. 27, l. 62. Thus, we disagree with Patent Owner that the '273 patent contrasts LHPDR execution units with those that operate in both high and low precision. PO Resp. 19, 23. For the same reasons, we assign little weight to Dr. Khatri's testimony on this issue. *See* Ex. 2051 ¶¶ 63–64, 72.

To the contrary, we agree with Petitioner that the '273 patent discloses at least one embodiment that performs operations other than LPHDR computations:

> [W]hile the obvious method of using the above [logarithmic number system (LNS)] operations is to do LPHDR arithmetic, the programmer also may consider selected values to be 12 bit two's complement binary numbers. MUL and DIV may be used to add and subtract such values . . . . *So besides doing LPHDR computations*, this digital embodiment using LNS can perform simple binary arithmetic on short signed integers.

Ex. 1001, col. 14, ll. 6–15 (emphasis added), *quoted in* Reply 4. Because the unit performs arithmetic operations other than those at low precision, this

16

IPR2021-00179
Patent 8,407,273 B2

example contradicts Patent Owner's argument that an LPHDR execution unit is "an execution unit that executes arithmetic operations *only* at low precision." PO Resp. 15 (emphasis added).

Patent Owner counters, "as Dr. Khatri explained, a [person of ordinary skill in the art] would not apply [the terms] 'precision' or 'dynamic range' to integers." Sur-Reply 4 (citing Ex. 1072, 116:5–118:3). In Patent Owner's view, a person of ordinary skill in the art "would understand that [Patent Owner's] proposed construction would *not* exclude execution units that can perform integer operations, like the LNS embodiment described in the patent specification." *Id.* at 4–5. Yet Patent Owner's proposed construction makes no distinction between integer or floating-point operations. PO Resp. 15. Rather, the construction refers to "arithmetic operations," generally. *See id.*

In its Sur-Reply, Patent Owner proposes an alternative construction—for the first time in this proceeding—that further defines "arithmetic operations" to be "*floating-point* arithmetic operations." Sur-Reply 5. But a sur-reply must only respond to arguments raised in the preceding brief. Patent Trial and Appeal Board Consolidated Trial Practice Guide (Nov. 2019), 73–74, *available at* https://www.uspto.gov/ TrialPracticeGuideConsolidated. Under 37 C.F.R. § 42.23(b), "respond" does not mean proceed in a new direction with a new approach. *See id.* at 74. "[A] sur-reply that raises a new issue or belatedly presents evidence may not be considered." *Id.* So, apart from noting that we do not see any reason to limit the LPHDR execution unit to only low precision operations— regardless of how those operations are defined—for the reasons discussed above, we do not further consider this new claim construction presented for the first time in the Sur-Reply. *See* Sur-Reply 5.

17

IPR2021-00179
Patent 8,407,273 B2

In sum, we agree with Petitioner that the recited "low precision high dynamic range (LPHDR) execution unit" does not exclude units that are capable of performing operations at low and full precision, and the degree of "low precision" is defined in limitation *1B2*. *See, e.g.*, Pet. 13–14, 30; Reply 2–3. To determine the patentability of the challenged claims, we need not further construe "low precision high dynamic range (LPHDR) execution unit." *See Nidec*, 868 F.3d at 1017.

C. Legal Standards

To prevail in its challenges to the patentability of claims 1–26, 28, 32–61, 63, and 67–70 of the '273 patent, Petitioner must demonstrate by a preponderance of the evidence that the claims are unpatentable. 35 U.S.C. § 316(e). "In an [*inter partes* review], the petitioner has the burden from the onset to show with particularity why the patent it challenges is unpatentable." *Harmonic Inc. v. Avid Tech., Inc.*, 815 F.3d 1356, 1363 (Fed. Cir. 2016). This burden of persuasion never shifts to Patent Owner. *Dynamic Drinkware, LLC v. Nat'l Graphics, Inc.*, 800 F.3d 1375, 1378 (Fed. Cir. 2015); *see also In re Magnum Oil Tools Int'l, Ltd.*, 829 F.3d 1364, 1376 (Fed. Cir. 2016) ("Where, as here, the only question presented is whether due consideration of the four *Graham* factors renders a claim or claims obvious, no burden shifts from the patent challenger to the patentee.").

A claim is unpatentable for obviousness if, to one of ordinary skill in the pertinent art, "the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007) (quoting 35 U.S.C. § 103(a) (2006)).

IPR2021-00179
Patent 8,407,273 B2

The question of obviousness is resolved on the basis of underlying factual determinations, including "the scope and content of the prior art"; "differences between the prior art and the claims at issue"; and "the level of ordinary skill in the pertinent art." *Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966). Additionally, objective indicia of nonobviousness, such as "commercial success, long felt but unsolved needs, failure of others, etc., might be utilized to give light to the circumstances surrounding the origin of the subject matter sought to be patented. As indicia of obviousness or nonobviousness, these inquiries may have relevancy." *Id.* When conducting an obviousness analysis, we consider a prior art reference "not only for what it expressly teaches, but also for what it fairly suggests." *Bradium Techs. LLC v. Iancu*, 923 F.3d 1032, 1049 (Fed. Cir. 2019) (citation omitted).

A patent claim "is not proved obvious merely by demonstrating that each of its elements was, independently, known in the prior art." *KSR*, 550 U.S. at 418. An obviousness determination requires finding "both 'that a skilled artisan would have been motivated to combine the teachings of the prior art references to achieve the claimed invention, and that the skilled artisan would have had a reasonable expectation of success in doing so.'" *Intelligent Bio-Sys., Inc. v. Illumina Cambridge Ltd.*, 821 F.3d 1359, 1367–68 (Fed. Cir. 2016) (citation omitted); *see KSR*, 550 U.S. at 418 (for an obviousness analysis, "it can be important to identify a reason that would have prompted a person of ordinary skill in the relevant field to combine the elements in the way the claimed new invention does"). Also, "[t]hough less common, in appropriate circumstances, a patent can be obvious in light of a single prior art reference if it would have been obvious to modify that reference to arrive at the patented invention." *Arendi S.A.R.L. v. Apple Inc.*, 832 F.3d 1355, 1361 (Fed. Cir. 2016).

IPR2021-00179
Patent 8,407,273 B2

"Although the *KSR* test is flexible, the Board 'must still be careful not to allow hindsight reconstruction of references . . . without any explanation as to *how* or *why* the references would be combined to produce the claimed invention.'" *TriVascular, Inc. v. Samuels*, 812 F.3d 1056, 1066 (Fed. Cir. 2016) (citation omitted). Further, an assertion of obviousness "cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness." *KSR*, 550 U.S. at 418 (quoting *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006)); *accord In re NuVasive, Inc.*, 842 F.3d 1376, 1383 (Fed. Cir. 2016) (stating that "conclusory statements" amount to an "insufficient articulation[] of motivation to combine"; "instead, the finding must be supported by a 'reasoned explanation'" (citation omitted)); *Magnum Oil*, 829 F.3d at 1380 ("To satisfy its burden of proving obviousness, a petitioner cannot employ mere conclusory statements. The petitioner must instead articulate specific reasoning, based on evidence of record, to support the legal conclusion of obviousness.").

D. *Obviousness Ground Based on Dockser (Claims 1, 2, 21–24, 26, and 28)*

1. *Dockser*

Dockser discloses performing floating-point operations with a floating-point processor having "selectable subprecision." Ex. 1007, code (57), ¶¶ 15, 17. "A floating-point representation of a number commonly includes a sign component, an exponent, and a mantissa. To find the value of a floating-point number, the mantissa is multiplied by a base (commonly 2 in computers) raised to the power of the exponent. The sign is applied to the resultant value." *Id.* ¶ 1. "The precision of the floating-point processor is defined by the number of bits used to represent the mantissa.

20

IPR2021-00179
Patent 8,407,273 B2

The more bits in the mantissa, the greater the precision." *Id.* ¶ 2. "The precision of [a] floating-point processor generally depends on the particular application. For example, the ANSI/IEEE-754 standard (commonly followed by modern computers) specifies a 32-bit single format having a 1-bit sign, an 8-bit exponent, and a 23-bit mantissa." *Id.* "Higher precision results in a higher accuracy, but commonly results in increased power consumption." *Id.* Dockser explains that the performance of floating-point operations also can be computationally inefficient because "[w]hile some applications may require [high] types of precision, other applications may not." *Id.* ¶ 3. For example, "some graphics applications may only require a 16-bit mantissa," such that "any accuracy beyond 16 bits of precision tends to result in unnecessary power consumption," but other applications may require "greater precision." *Id.* Accordingly, there was "a need in the art for a floating-point processor in which the reduced precision, or subprecision, of the floating-point format is selectable." *Id.*

IPR2021-00179
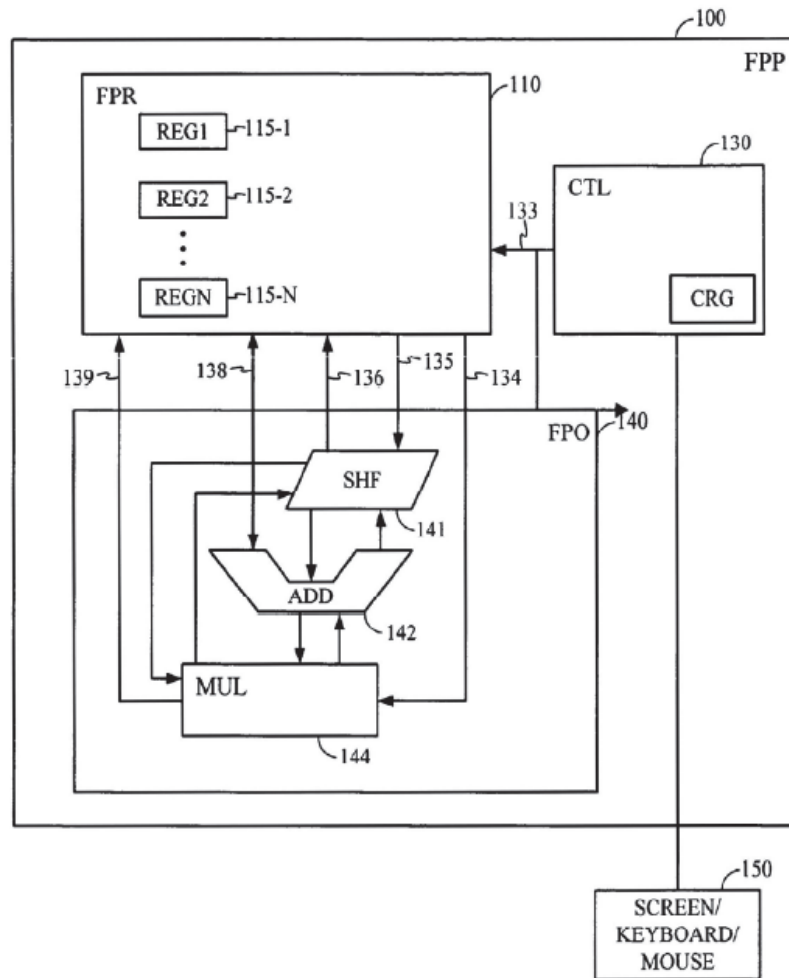Patent 8,407,273 B2

Figure 1 of Dockser is reproduced below.

FIG. 1

Figure 1 depicts floating-point processor (FPP) 100 including floating-point register file (FPR) 110 for storing floating-point numbers, floating-point controller (CTL) 130 "used to select the subprecision of the floating-point operations using a control signal 133," and floating-point mathematical operator (FPO) 140 with components "configured to perform the floating-point operations," such as floating-point adder (ADD) 142 and floating-point multiplier (MUL) 144. *Id.* ¶¶ 15, 18, 19.

22

IPR2021-00179
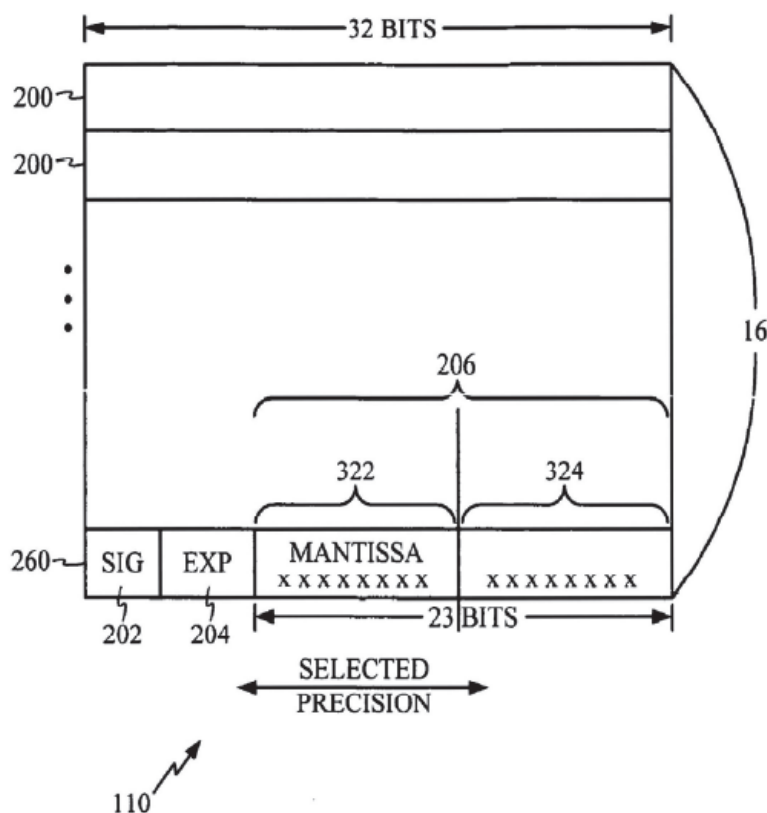Patent 8,407,273 B2

Figure 2 of Dockser is reproduced below.

FIG. 2

Figure 2 depicts an exemplary data structure for floating-point register file 110 including 16 addressable register locations 200, each "configured to store a 32-bit binary floating-point number" as "a 1-bit sign 202, an 8-bit exponent 204, and a [23-bit] fraction 206." *Id.* ¶ 17.

"[F]or each instruction of a requested floating-point operation, the relevant computational unit . . . receive[s] from the floating-point register file 110 one or more operands stored in one or more of the register locations" and executes the instruction "at the subprecision selected by the floating-point controller 130." *Id.* ¶¶ 23–24. The precision of the floating-point operation can be reduced by "caus[ing] power to be removed from the floating-point register elements for the excess bits of the fraction that are not required to meet the precision specified by the subprecision

IPR2021-00179
Patent 8,407,273 B2

select bits" written to the control register. *Id.* ¶¶ 6, 25–26. For example, "if each location in the floating-point register file contains a 23-bit fraction, and the subprecision required for the floating-point operation is 10-bits, only the 9 commonly significant bits (MSBs) of the fraction are required; the hidden or integer bit makes the tenth." *Id.* ¶ 26. "Power can be removed from the floating-point register elements for the remaining 14 fraction bits." *Id.*

Alternatively, power can be removed in elements of "the logic in the floating-point operator 140 that remains unused as a result of the subprecision selected." *Id.* ¶¶ 7, 27, 29, 32, Fig. 2 (depicting mantissa fraction 206 as having portion 322 for powered bits and portion 324 for unpowered bits). Figures 3A and 3B of Dockser show such removal of power to the floating-point operator logic for a floating-point addition and floating-point multiplication operation, respectively. Figure 3B of Dockser is reproduced below.
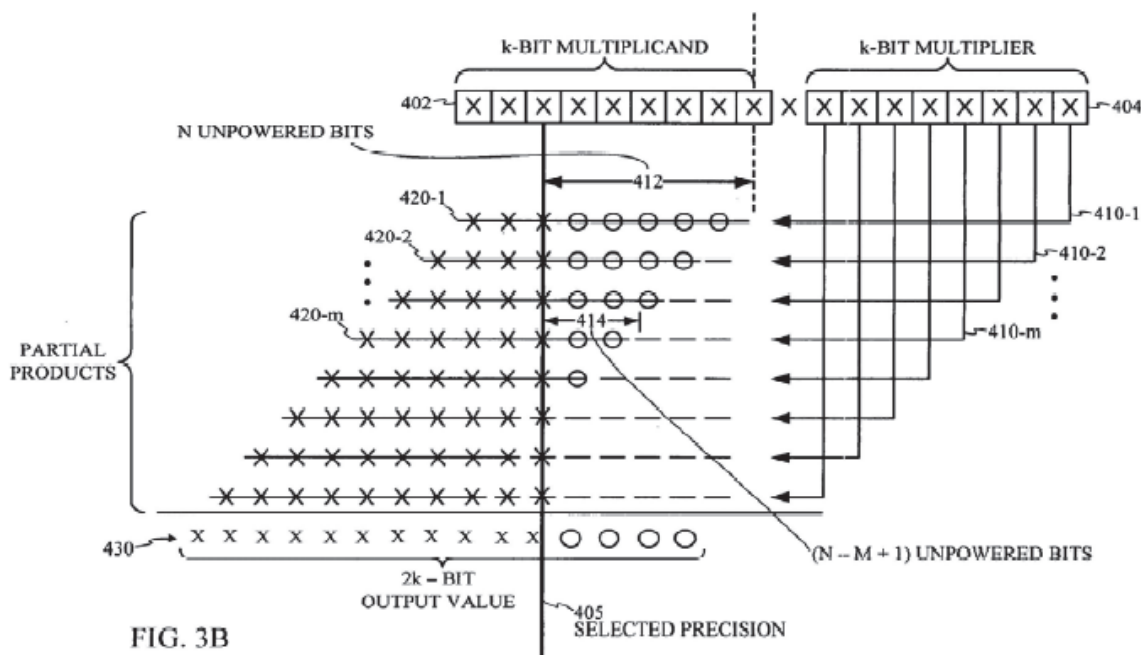

FIG. 3B

Figure 3B depicts k-bit multiplicand 402 and k-bit multiplier 404 to be multiplied together "using a shift-and-add technique," where the

24

IPR2021-00179
Patent 8,407,273 B2

multiplication occurs in stages 410-1 through 410-m. *Id.* ¶¶ 30–31.

A partial product 420-i is generated for every bit in multiplier 404 at

corresponding stage 410-i and then left-shifted "as a function of the

multiplier bit with which it is associated, after which the operation moves on

to the next stage." *Id.* ¶ 31. The partial products are eventually added

together to generate output value 430. *Id.* "[P]ower may be removed from

the logic used to implement the stages to the right of the line 405" indicating

the selected subprecision. *Id.* ¶¶ 32–33.

2.   Claim 1

Petitioner explains in detail how Dockser teaches or suggests every

limitation of claim 1, relying on the testimony of Mr. Goodin as support.

See Pet. 8–36; Ex. 1003 ¶¶ 204–310.

Claim 1 recites a device comprising "at least one first low precision

high dynamic range (LPHDR) execution unit adapted to execute a first

operation on a first input signal representing a first numerical value to

produce a first output signal representing a second numerical value."

Petitioner asserts that Dockser's floating-point processor (FPP) is a "low

precision high dynamic range (LPHDR) execution unit."[2] Pet. 12–14.

According to Petitioner, Dockser's FPP is incorporated in a computing

system, which is the claimed "device" comprising the FPP. *Id.* at 12 (citing

Ex. 1003 ¶ 204; Ex. 1007 ¶¶ 1, 15, 35–36). Petitioner argues that "[w]ithin

---

[2] Petitioner also provides an "alternative mapping" where "the floating-point
operator (FPO) inside Dockser's FPP" constitutes an LPHDR execution unit.
Pet. 13, 18–20, 35. We need not evaluate those arguments because we
determine that Petitioner has made a sufficient showing based on the full
trial record that the FPP is an LPHDR execution unit.

IPR2021-00179
Patent 8,407,273 B2

Dockser's computing system, the FPP is implemented in a 'hardware component[]' such as an [application specific integrated circuit (ASIC)] or [field-programmable gate array (FPGA)], which is also a claimed 'device' comprising the FPP." *Id.* (citing Ex. 1003 ¶ 206; Ex. 1007 ¶ 35). We agree with Petitioner.

a) *"High Dynamic Range"*

With respect to the "high dynamic range" aspect of the LPHDR execution unit, the claim requires that "the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/65,000 through 65,000." Petitioner asserts that Dockser's FPP "uses an 8-bit floating-point exponent . . . that provides an even higher dynamic range" than the 6-bit floating-point exponent disclosed in the '273 patent. Pet. 14 (citing Ex. 1007 ¶ 17; Ex. 1001, col. 14, ll. 53–61). Petitioner's assertion, which is not disputed by Patent Owner, is adequately supported by the record.

For example, Dockser's register location 200 stores a 32-bit binary floating-point number, in an IEEE-754 32-bit single format, with 8-bit exponent 204. Ex. 1007 ¶¶ 2, 17, *cited in* Pet. 19. According to Mr. Goodin, the dynamic range of normal IEEE-754 operands would be from around 2^{-126}, which is much smaller than 1/65,000, to around 2^{127}, which is much larger than 65,000. Ex. 1003 ¶ 241, *cited in* Pet. 19. We credit Mr. Goodin's testimony on this point because it is consistent with Dockser. *See id.*; Ex. 1007 ¶¶ 2, 16–17, 24. Thus, Petitioner has shown that Dockser's FPP meets the dynamic range limitation recited in claim 1.

26

IPR2021-00179
Patent 8,407,273 B2

b)   *"A First Operation"*

Claim 1 recites that the LPHDR execution unit is "adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value." Petitioner argues that Dockser's FPP is adapted to execute a "first operation" (e.g., "reduced-precision multiplication") on a "first input signal representing a first numerical value" (i.e., input electrical signal representing an operand received at the registers) to produce a "first output signal representing a second numerical value" (i.e., output electrical signal representing an operand sent to a register and then main memory). Pet. 16–18.

Patent Owner does not present an argument to specifically address this argument and evidence. *See* PO Resp.; Sur-Reply.

Based on our review of the totality of the evidence in this case, we agree that Dockser's FPP is adapted to execute the recited first operation. Dockser's FPP receives the operands as input values. Ex. 1007, Fig. 1. Dockser's FPP performs operations on the inputs via the FPP's data paths 134–139 and components 140–144. *Id.* We agree with Petitioner that Dockser teaches that the operand data to the FPP is a "first input signal representing a first numerical value" because it represents the operand number. *Id.* ¶ 17. The FPP executes the operation and produces an "output value," or "output number," which is represented by "output bits." *Id.* ¶ 34. In this way, that value is the recited "second numerical value." *See* Pet. 16–18. We credit Mr. Goodin's testimony on this point. Ex. 1003 ¶¶ 232–235.

We also credit Mr. Goodin's analysis showing that Dockser's computing circuits "move" data between components, such as from memory

27

IPR2021-00179
Patent 8,407,273 B2

to the processor, via electrical signals. *See id.* ¶ 229 (citing Ex. 1014, col. 1, ll. 36–37). On this issue, the record provides adequate support for Mr. Goodin's testimony. *See* Ex. 1007 ¶¶ 16–17; Ex. 1014, col. 1, ll. 36–37. Thus, Petitioner has shown that signals represent the recited values.

Considering all of the arguments and evidence in this case, Petitioner has sufficiently shown that Dockser teaches that the LPHDR execution unit is "adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value," as recited in claim 1. Pet. 14–18.

c)   *"Low Precision"*

As for the "low precision" aspect of the LPHDR execution unit, Patent Owner argues that Dockser's FPP is not an LPHDR execution unit because the FPP is capable of both low and full precision operations. *See* PO Resp. 15, 20–24. We disagree with Patent Owner's argument and underlying construction for the reasons discussed in Section II.B. Instead, we agree with Petitioner that (1) the recited "low precision high dynamic range (LPHDR) execution unit" does not exclude units that perform operations at low and full precision, and (2) the degree of "low precision" required by the claim is defined in limitation *1B2*. *See, e.g.*, Pet. 13–14, 30; Reply 2–3.

d)   *Limitation 1B2*

Under the interpretation discussed in Section II.B, Petitioner has shown that Dockser teaches the recited LPHDR execution unit. In particular, Petitioner argues that the FPP is "low precision" because "'the precision' of operations in the FPP is 'reduced'" and because the FPP

28

IPR2021-00179
Patent 8,407,273 B2

"operates with the minimum imprecision" required by limitation *1B2*.

Pet. 13–14 (quoting Ex. 1007 ¶ 14). Limitation *1B2* recites (emphasis added):

> for at least X=5% of the possible valid inputs to the first operation, the *statistical mean*, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an *exact mathematical calculation* of the first operation on the numerical values of that same input.

We first analyze the "statistical mean" limitation before turning to the "exact mathematical calculation" limitation.

(1) "Statistical Mean"

Petitioner argues that a person of ordinary skill in the art would have understood the "statistical mean" limitation

> in the context of the '273 patent's stated intent to claim not only "repeatable" deterministic embodiments like digital circuits that always produce the same output when repeating an operation on the same input, but also analog embodiments that are non-deterministic because they "introduce noise into their computations, so the computations are not repeatable."

Pet. 20–21 (quoting Ex. 1001, col. 4, ll. 7–13).

In Petitioner's view, for "non-deterministic embodiments," the statistical mean would be the average of "the different outputs produced by the same operation on the same input." *Id*. at 21. Petitioner argues that, for "deterministic digital embodiments" like Dockser, the statistical mean is "the same as the numerical value of the first output signal for any individual execution of the first operation on each specific input, because that output is always the same for any specific input." *Id*. In Petitioner's view, repeatedly

29

IPR2021-00179
Patent 8,407,273 B2

executing a multiplication operation using Dockser's floating-point multiplier "on the same input (*i.e.*, pair of operands) with the same precision level yields the same result for every execution; therefore, the statistical mean of the outputs is the same as the output for any single execution." *Id.* at 22.

Patent Owner does not dispute this reasoning. *See* PO Resp.; Sur-Reply.

Based on the totality of the evidence, Petitioner's analysis of "deterministic digital embodiments" to address the "statistical mean" limitation is adequately supported by the record. Pet. 21–22. We are persuaded by Petitioner's reasoning that the '273 patent's written description has both analog and digital embodiments. *See, e.g.*, Ex. 1001, col. 4, ll. 7–13, col. 14, ll. 16–61, col. 17, ll. 10–14. Dockser uses "a conventional floating-point multiplier." Ex. 1007 ¶ 20. Petitioner adequately supports its assertion that this is a deterministic digital circuit. Pet. 22 (citing Ex. 1003 ¶ 250; Ex. 1011, col. 1, ll. 40–42). And we credit Mr. Goodin's testimony that, in such a digital circuit, a floating-point multiplier operating on the same input with the same precision level yields the same result for every execution, and thus, the statistical mean is the same as the output for any single execution. Ex. 1003 ¶ 251, *cited in* Pet. 22.

(2) "Exact Mathematical Calculation"

As for the "exact mathematical calculation" limitation, Petitioner argues that because Dockser performs a reduced precision multiplication, the result of the operation differs from what would be the exact mathematical result of the operation: "the (>32-bit) product that would result if the pair of

30

IPR2021-00179
Patent 8,407,273 B2

input 32-bit operands were multiplied without reducing precision."

Pet. 22–23 (emphasis omitted) (citing Ex. 1003 ¶ 253; Pet. App'x I.B).

(3) Y Relative Error for X Valid Inputs

According to Petitioner, "Dockser discloses two precision-reducing techniques that can be used separately or together." Pet. 25 (citing Ex. 1007 ¶¶ 4–7; Ex. 1003 ¶¶ 264–270). One technique removes power from storage elements in the FPP registers that correspond to the excess, dropped mantissa bits, which drops bits from the operands. *Id.* The other technique "remov[es] power from elements within the multiplier logic that computes the product of the operand mantissas." *Id.* Petitioner asserts that both techniques individually and in combination teach limitation *1B2*.

We determine that Petitioner has made a sufficient showing that Dockser meets the recited requirements for low precision under the register bit-dropping rationale. Patent Owner does not present arguments specific to this technique. *See* PO Resp.; Sur-Reply. We need not evaluate Petitioner's alternative rationale based on the multiplier logic. Our reasoning follows.

(4) Register Bit-Dropping

Petitioner explains how Dockser teaches that "for at least X=5% of the possible valid inputs to the first operation," the statistical mean of the results of executing the first operation "differs by at least Y=0.05%" from the result of the exact mathematical calculation. Pet. 23–36. In Petitioner's view, the "possible valid inputs" in Dockser are "the set of possible normal IEEE-754 32-bit single-format numbers forming pairs of operands in input signals to the execution unit that can be multiplied together to produce an output representing a numerical value (rather than, *e.g.*, an

31

IPR2021-00179
Patent 8,407,273 B2

overflow/underflow exception)." *Id.* at 23. Petitioner argues that Dockser's

FPP operates at a precision level meeting the claimed X and Y percentages

for such input pairs. *Id.* at 23–24.

We agree. Dockser retains only some of the bits of a mantissa

fraction: in one example, the nine most-significant bits of the mantissa

fraction are retained, and the other 14 excess bits are dropped. Ex. 1007

¶¶ 25–28, *cited in* Pet. 24; Ex. 1003 ¶ 263. The outputs of Dockser's

unpowered storage elements are tied to zero voltage. Ex. 1003 ¶ 278;

Ex. 1007 ¶¶ 26 (discussing removing power from the register elements for

the excess bits), 29 (discussing "unpowered bits"). With support from

Mr. Goodin and corroborating evidence, Petitioner shows that tying the

outputs of the unpowered storage elements for the 14 "excess" bits to zero

voltage would make those bits zeroes. Pet. 28 (citing Ex. 1003 ¶¶ 278–280).

Multiplying two operands with excess bits dropped (representing

zero) reduces the output's precision. Ex. 1007 ¶¶ 24–26. Petitioner

provides a representative example described by Mr. Goodin:

> the 23-bit sequence 01100000111100111001110 (representing a
> mantissa of 1.3787171840667724609375) would become
> 01100000100000000000000 (representing a mantissa of
> 1.376953125) by zeroing the 14 right-most bits, and
> 01000100110111000001000 (representing a mantissa of
> 1.26898288726806640625) would become
> 01000100100000000000000 (representing a mantissa of
> 1.267578125).

Pet. 28–29 (citing Ex. 1003 ¶ 282). So the decrease in precision depends on

the number of mantissa bits dropped. *See id.*; *see also* Ex. 1003 ¶ 287. And

we agree with Petitioner that Dockser suggests that any desired number of

mantissa bits can be dropped. Pet. 30. For example, Dockser describes how

IPR2021-00179
Patent 8,407,273 B2

the precision is selectable. *See* Ex. 1007 ¶¶ 14–15, 17 (describing selectable reduced precision), *discussed in* Pet. 4.

Petitioner asserts that Dockser's register bit-dropping technique meets limitation *1B2* in two ways. Pet. 26–32. First, Petitioner presents results of a software demonstration to show that Dockser's 9 mantissa bits meets limitation *1B2*. *Id.* at 30–31. Second, Petitioner uses a pencil-and-paper algebraic analysis to show that retaining 9 mantissa bits meets limitation *1B2*. *Id.* at 31–32.

As for the software demonstration, Petitioner states that "[g]iven the massive number of possible inputs to Dockser's FPP (including over 70 trillion possible pairs of normal IEEE-754 single-format mantissas)," a person of ordinary skill in the art "would have performed Dockser's FPP operation in software to determine the fraction X of all possible valid inputs that produce at least the claimed relative error Y when a given number of mantissa bits are dropped." *Id.* at 30–31. Petitioner states that Mr. Goodin wrote a program to perform reduced precision multiplication retaining 9 mantissa bits and dropping the 14 excess bits as in Dockser that tested all possible valid mantissa pairs and "produce[d] at least Y=0.05% relative error for 92.1% of possible valid inputs (greater than X=5%)." *Id.* at 30–31, 68–70 (citing Ex. 1003 ¶¶ 290–292, 441–455, 495–500).

Patent Owner did not rebut this evidence. *See* PO Resp.; Sur-Reply.

We are persuaded by Petitioner's assertions about the recited X and Y percentages, and we credit Mr. Goodin's testimony on this issue, which is supported by the record. *See* Ex. 1003 ¶¶ 441–455.

33

IPR2021-00179
Patent 8,407,273 B2

In particular, Mr. Goodin summarizes the results of the software program in the table below. *Id.* ¶ 452.

Retained Fraction Bits	Claimed Y%	Percentage of Tested Pairs Meeting Claimed Y%	Adjusted "Dockser's X%"
9	≥ 0.05%	92.628313%	≥ 92.14333%
	≥ 0.10%	70.494075%	≥ 70.12498%
	≥ 0.15%	39.118049%	≥ 38.91323%
	≥ 0.20%	14.674747%	≥ 14.59791%

5	≥ 0.05%	99.971764%	≥ 99.44834%
	≥ 0.10%	99.886981%	≥ 99.36400%
	≥ 0.15%	99.976929%	≥ 99.22327%
	≥ 0.20%	99.547206%	≥ 99.02600%

The table shows the percentage of tested pairs that meet the claimed Y percentage as output by the software program. *Id.* It also shows a "Dockser's X" percentage, which is adjusted to exclude possible overflow and underflow pairs, that meets the claimed requirements. *Id.* The X value is calculated by multiplying the percentage output by the software program by an adjustment ratio. *Id.* The adjustment ratio is described in Appendix I.E of Mr. Goodin's declaration. *Id.* Mr. Goodin also explains that, if "the 'exact mathematical calculation' of multiplication of two single-format floating-point operands is the product rounded or truncated to single format (having a 23-bit-fraction mantissa), the claims are still met." *Id.* ¶ 455. To show this, Mr. Goodin wrote an alternative version of the program under this assumption, which also gives results that meet the claim limitations. *Id.* After review of the arguments and evidence on these issues, we credit Mr. Goodin's software-based analysis. *Id.* ¶¶ 441–455.

As for the algebraic analysis, Petitioner argues that a person of ordinary skill in the art "would also have understood algebraically that

IPR2021-00179
Patent 8,407,273 B2

Dockser's register bit-dropping technique meets [limitation *1B2*], by examining the absolute *minimum* relative error produced by zeroing certain mantissa bit positions." Pet. 31–32. Petitioner explains how over 12% of possible input operands "have a zero as their most-significant (leftmost) mantissa fraction bit and ones as their tenth and eleventh fraction bits," such that retaining only 9 mantissa bits of operands would result in "*every* input in that 12% produc[ing] at *minimum* 0.097% relative error." *Id.* at 31–32, 71–76 (citing Ex. 1003 ¶¶ 293–294, 456–475).

Patent Owner did not rebut this evidence. *See* PO Resp.; Sur-Reply.

We are persuaded by Petitioner's algebraic method showing the recited X and Y percentages, and credit Mr. Goodin's testimony on this issue. *See* Ex. 1003 ¶¶ 456–475. Under this reasoning, "for the 12.5% of all possible M_A [mantissa] values with zero in the first fraction bit and ones in the $(K + 1)^{th}$ and $(K + 2)^{th}$ fraction bits," the following inequality holds "when Dockser's register bit-dropping truncates the operands to K fraction bits":

$$Y \geq \frac{\left(2^{-(K+1)} + 2^{-(K+2)}\right)}{(1.5 - 2^{-23})} \times 100\% \qquad \textbf{\textit{(Equation C)}}$$

Id. ¶ 471. The table below shows the results of evaluating Equation C with K=9, K=8, K=7, and K=5 retained fraction bits as the selected precision level. *Id.* ¶ 475.

Retained Fraction Bits (K)	Equation C: Minimum Y for $X \geq 12\%$	Equation D: Minimum Y for $X \geq 6\%$	Meets X/Y Percentages Recited by Claims:
9	≥ 0.0976%	≥ 0.1171%	1, 11-12, 33, 36, 46-47, 68
8	≥ 0.1953%	≥ 0.2343%	All above plus 13-16, 48-51
7	≥ 0.3906%	≥ 0.4687%	All above plus 17, 52
5	≥ 1.5625%	≥ 1.8750%	All above

IPR2021-00179
Patent 8,407,273 B2

The table above shows that the values of K meet the '273 patent's claimed

minimum X and Y. *Id.* After review of the arguments and evidence on

these issues, we credit Mr. Goodin's algebraic analysis. *Id.* ¶¶ 456–475.

e)   Objective Indicia of Nonobviousness

In addition to its argument based on the interpretation of "low

precision high dynamic range (LPHDR) execution unit," Patent Owner

argues that objective indicia of nonobviousness demonstrate that the

challenged claims would not have been obvious to a person of ordinary skill

in the art. PO Resp. 46–72. Because the parties largely refer to the

challenged claims collectively in their arguments regarding objective indicia,

we do so as well and now address those arguments as applied to all of the

claims challenged as obvious over Dockser alone or the combination of

Dockser and Tong, i.e., claims 1, 2, 21–24, 26, 28, 32, and 33, which we

refer to as the "Dockser/Tong-challenged claims."[3] *See id.*; Reply 19–30;

Sur-Reply 19–26.

"In order to accord substantial weight to secondary considerations in

an obviousness analysis, the evidence of secondary considerations must have

a nexus to the claims, *i.e.*, there must be a legally and factually sufficient

---

[3] As explained below, we determine that Petitioner has not proven that
certain claims are unpatentable over the combinations of Dockser and
MacMillan, and Dockser, Tong, and MacMillan, because Petitioner has not
sufficiently shown a reason to combine with MacMillan. *See infra*
Sections II.F, II.G. Accordingly, we need not assess objective indicia of
nonobviousness with respect to those claims. *See Mylan Pharms. Inc. v.
Res. Corp. Techs., Inc.*, 914 F.3d 1366, 1376 (Fed. Cir. 2019) ("Because we
agree with the Board that Appellants failed to establish a motivation to
modify compound 3I, we need not reach Appellants' arguments regarding
objective indicia.").

IPR2021-00179
Patent 8,407,273 B2

connection between the evidence and the patented invention." *Fox Factory, Inc. v. SRAM, LLC*, 944 F.3d 1366, 1373 (Fed. Cir. 2019) (citations and internal quotation marks omitted). "The patentee bears the burden of showing that a nexus exists." *Id.* "To determine whether the patentee has met that burden, we consider the correspondence between the objective evidence and the claim scope." *Id.* A patentee is entitled to a rebuttable presumption of nexus "when the patentee shows that the asserted objective evidence is tied to a specific product and that product 'embodies the claimed features, and is coextensive with them.'" *Id.* (citation omitted). Whether a rebuttable presumption of nexus arises "turns on the nature of the claims and the specific facts." *Teva Pharms. Int'l GmbH v. Eli Lilly & Co.*, 8 F.4th 1349, 1362 (Fed. Cir. 2021). "As part of the presumption analysis, the fact finder must consider the unclaimed features of the stated products to determine their level of significance and their impact on the correspondence between the claim and the products." *Quanergy Sys., Inc. v. Velodyne Lidar USA, Inc.*, 24 F.4th 1406, 1418 (Fed. Cir. 2022). When, "for example, the patented invention is only a small component of the product tied to the objective evidence, there is no presumption of nexus." *Henny Penny Corp. v. Frymaster LLC*, 938 F.3d 1324, 1333 (Fed. Cir. 2019). Absent a presumption of nexus, a patent owner may "prove nexus by showing that the evidence of secondary considerations is the 'direct result of the unique characteristics of the claimed invention.'" *Fox Factory*, 944 F.3d at 1373–74 (citation omitted).

"Where the offered secondary consideration actually results from something other than what is both claimed and *novel* in the claim, there is no nexus to the merits of the claimed invention," meaning that "there must be a nexus to some aspect of the claim not already in the prior art." *In re Kao*,

IPR2021-00179
Patent 8,407,273 B2

639 F.3d 1057, 1068–69 (Fed. Cir. 2011). On the other hand, there is no
requirement that "objective evidence must be tied exclusively to claim
elements that are not disclosed in a particular prior art reference in order for
that evidence to carry substantial weight." *WBIP, LLC v. Kohler Co.*,
829 F.3d 1317, 1331 (Fed. Cir. 2016). A patent owner may show, for
example, "that it is the claimed combination as a whole that serves as a
nexus for the objective evidence; proof of nexus is not limited to only when
objective evidence is tied to the supposedly 'new' feature(s)." *Id.* at 1330.
Ultimately, the fact finder must weigh the objective indicia evidence
presented in the context of whether the claimed invention as a whole would
have been obvious to an ordinarily skilled artisan. *Id.* at 1331–32;
Lectrosonics, Inc. v Zaxcom, Inc., IPR2018-01129, Paper 33 at 33 (PTAB
Jan. 24, 2020) (precedential).

(1) Industry Skepticism

"If industry participants or skilled artisans are skeptical about whether
or how a problem could be solved or the workability of the claimed solution,
it favors non-obviousness. Doubt or disbelief by skilled artisans regarding
the likely success of a combination or solution weighs against the notion that
one would combine elements in references to achieve the claimed
invention." *WBIP*, 829 F.3d at 1335. As evidence of alleged industry
skepticism, Patent Owner points to various statements made by employees
of Petitioner after meeting with Dr. Bates (the named inventor of the
'273 patent) in 2010. PO Resp. 47–51. According to Patent Owner,
Dr. Bates explained that floating point units programmed to perform

IPR2021-00179
Patent 8,407,273 B2

███

███

██████████████████████████████ *Id.* at 48–51 (citing Exs. 2009, 2027, 2028, 2030). Patent Owner contends that the cited statements from Petitioner's employees "express skepticism about utility of the low-precision arithmetic performed by the large numbers of low-precision execution units,"[4] which is a "key feature" of the challenged claims, and have a nexus to the challenged claims because they were made in response to Dr. Bates's presentations. *Id.* at 51.

Petitioner argues that Patent Owner's evidence is irrelevant because it is dated after the earliest possible effective filing date of the '273 patent (June 19, 2009). Reply 22. We disagree that evidence of industry skepticism must itself be dated before the relevant invention. *See WBIP*, 829 F.3d at 1335–36 (considering evidence of skepticism dated after the invention of the patent at issue); *Pressure Prods. Med. Supplies, Inc. v. Greatbatch Ltd.*, 599 F.3d 1308, 1319 (Fed. Cir. 2010) (same); Sur-Reply 19–20.

---

[4] Claim 1 recites "at least one" LPHDR execution unit and thus is satisfied by only a single LPHDR execution unit. The only challenged claims of the '273 patent that arguably require a "large number[]" of LPHDR execution units are challenged in Petitioner's asserted grounds based on combinations with MacMillan. *See* PO Resp. 51. For example, claim 7 recites that "the at least one first LPHDR execution unit comprises at least one hundred LPHDR execution units," and claim 8 recites that "the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide."

IPR2021-00179
Patent 8,407,273 B2

We have reviewed the substance of the statements cited by Patent Owner, though, and are not persuaded that they are probative of nonobviousness. The statements by Petitioner's engineers relate to whether applications using "approximate" arithmetic would be commercially valuable; they do not show skepticism as to whether or how a problem could be solved or whether such applications would *work*. *See WBIP*, 829 F.3d at 1335. For example, Tom Dean stated that

Ex. 2009, 1 (emphasis added).

Ex. 2029, 1 (emphasis added).

Ex. 2028, 1 (emphasis added).

Ex. 2030, 2. Thus, the cited comments appear to indicate that floating point units performing "approximate" arithmetic would only be commercially valuable for a limited set of applications, rather than expressing skepticism as to whether such applications would work. Patent Owner also has not explained sufficiently why the fact that "approximate" arithmetic was considered useful for some

40

IPR2021-00179
Patent 8,407,273 B2

applications, but not others, amounts to industry skepticism over the claimed

invention as a whole (which is not limited to a particular application). We

find that Patent Owner's evidence of alleged industry skepticism does not

weigh in favor of nonobviousness of the Dockser/Tong-challenged claims.

(2) Industry Praise

"Evidence that the industry praised a claimed invention or a product

which embodies the patent claims weighs against an assertion that the same

claim would have been obvious. Industry participants, especially

competitors, are not likely to praise an obvious advance over the known art."

WBIP, 829 F.3d at 1334. Patent Owner points to four statements of alleged

industry praise from "leading figures in computer science" and argues that

they have a nexus to the challenged claims because they relate to Dr. Bates

and his presentations to Petitioner in 2010, 2011, and 2013. PO Resp.

51–52.

Patent Owner cites Dr. Bates's notes from ████████████████

██

████████████████████████████████   *See id.* at 51; Ex. 2050. This single

line in the notes does not mention any specific ideas or features that

allegedly were praised, or why or for what they were being praised. Patent

Owner argues that there is "no requirement that praise identify exactly the

features being praised." Sur-Reply 20. But evidence of industry praise must

be for "the claimed invention," and the notes only state that ███████████

███████████████████████   *See* Ex. 2050; *Apple Inc. v. Samsung Elecs. Co.*,

839 F.3d 1034, 1053 (Fed. Cir. 2016) (en banc); *WBIP*, 829 F.3d at 1334.

The notes also have less probative value because they were written by

Dr. Bates based on his subjective recollection of Mr. Dean's reaction at the

meeting, rather than, for example, being prepared by a third party or Mr. Dean himself. Ex. 2052 ¶ 7.

Patent Owner also cites three 2011 emails sent by employees of Petitioner to Dr. Bates. PO Resp. 51–52. Andrew Ng wrote the following to Dr. Bates:

Ex. 2033, 1. Astro Teller wrote that   and Johnny Chen wrote,   *See* Exs. 2031, 2032. We view these emails as more consistent with Petitioner's view as "merely cordial statements made to Dr. Bates," rather than evidence of industry praising the inventions of the Dockser/Tong-challenged claims as Patent Owner contends. *See* Reply 24. Further, they are statements by employees of Petitioner only (not any other industry participant or industry group) and do not mention any specific ideas or features from Dr. Bates's presentations that were being praised. We find that Patent Owner's evidence of alleged industry praise is too ambiguous and high-level to be probative of nonobviousness.

IPR2021-00179
Patent 8,407,273 B2

(3) Unexpected Results

"To be particularly probative, evidence of unexpected results must establish that there is a difference between the results obtained and those of the closest prior art, and that the difference would not have been expected by one of ordinary skill in the art at the time of the invention." *Bristol-Meyers Squibb Co. v. Teva Pharms. USA, Inc.*, 752 F.3d 967, 977 (Fed. Cir. 2014). Also, "[w]hen assessing unexpected properties, . . . we must evaluate the significance and 'kind' of expected results along with the unexpected results." *Id.* As evidence of alleged unexpected results, Patent Owner points to a September 2013 presentation that Dr. Bates gave to Petitioner showing

███

████████████████████████ PO Resp. 52–56 (citing Ex. 2005). Patent Owner argues that the unexpected nature of the results is "confirmed by skepticism expressed by" David Patterson, Ph.D. (an employee of Petitioner and "acclaimed expert in the field of computer science and computer architecture" according to Patent Owner), and "further supported by Dockser and Tong," which teach that higher precision is "generally" needed. *Id.* at 55–56 (citing Ex. 1007 ¶ 3; Ex. 1008, 280; Ex. 2012; Ex. 2051 ¶¶ 124–125).

The cited evidence, however, does not establish a difference between the alleged results and the closest prior art. Dr. Bates's presentation includes ████████████████████████████████████ Ex. 2005, 11; *see* PO Resp. 54. In this case, however, the closest prior art is Dockser and Tong, both of which teach reduced precision arithmetic and, importantly, show that high performance with reduced precision was *expected* for certain applications. *See* Reply 25; Ex. 1007 ¶ 3 ("[F]or certain applications, . . . a reduced precision, may be acceptable, and for other applications, . . . a greater precision may be needed."); Ex. 1008, 273 ("[M]any programs which

43

IPR2021-00179
Patent 8,407,273 B2

manipulate human sensory inputs, e.g., speech and image recognition, suffer no loss of accuracy with reduced bitwidth in the mantissa or exponent.").") Patent Owner contends that Dockser and Tong are not appropriate for comparison because they do not "provide computing scale results" and have an "entirely different focus on power savings." Sur-Reply 21. We disagree. As explained above, we find that Dockser teaches the claim 1 LPHDR execution unit performing the recited reduced precision arithmetic. *See supra* Section II.D.2. Patent Owner's only comparison, however, is to a CPU without any reduced precision capability. *See* PO Resp. 54. We find that Patent Owner's evidence of alleged unexpected results does not weigh in favor of nonobviousness of the Dockser/Tong-challenged claims because it does not establish a difference between the results and the closest prior art and does not establish that such a difference would have been unexpected by a person of ordinary skill in the art.

(4) Copying and Commercial Success

"Copying may . . . be another form of flattering praise for inventive features," as "[t]he fact that a competitor copied technology suggests it would not have been obvious." *WBIP*, 829 F.3d at 1336 (citation omitted). Copying "requires evidence of efforts to replicate a specific product." *Wyers v. Master Lock Co.*, 616 F.3d 1231, 1246 (Fed. Cir. 2010). "This may be demonstrated either through internal documents; direct evidence such as disassembling a patented prototype, photographing its features, and using the photograph as a blueprint to build a virtually identical replica; or access to, and substantial similarity to, the patented product (as opposed to the patent)." *Iron Grip Barbell Co. v. USA Sports, Inc.*, 392 F.3d 1317, 1325 (Fed. Cir. 2004) (internal citations omitted). However, "a showing of

44

IPR2021-00179
Patent 8,407,273 B2

copying is only equivocal evidence of non-obviousness in the absence of

more compelling objective indicia of other secondary considerations."

Ecolochem, Inc. v. S. Cal. Edison Co., 227 F.3d 1361, 1380 (Fed. Cir. 2000);

see also In re GPAC Inc., 57 F.3d 1573, 1580 (Fed. Cir. 1995) ("[M]ore than

the mere fact of copying by an accused infringer is needed to make that

action significant to a determination of the obviousness issue." (citation

omitted)).

Commercial success is typically shown with evidence of "significant

sales in a relevant market." *Ormco Corp. v. Align Tech., Inc.*, 463 F.3d

1299, 1312 (Fed. Cir. 2006) (citation omitted). "When a patentee can

demonstrate commercial success, usually shown by significant sales in a

relevant market, and that the successful product is the invention disclosed

and claimed in the patent, it is presumed that the commercial success is due

to the patented invention." *J.T. Eaton & Co. v. Atlantic Paste & Glue Co.*,

106 F.3d 1563, 1571 (Fed. Cir. 1997).

Patent Owner argues that after Dr. Bates made his presentations to

Petitioner in 2010–2014, Petitioner copied Dr. Bates's invention in its

Tensor Processing Unit (TPU) v2 and v3 products released in 2017–2018,

which "use execution units designed to perform operations using a

low-precision 'floating point format called bfloat16' in order to achieve

increased parallelism and scale." PO Resp. 56–57 (citing Ex. 2010;

Ex. 2011, 9; Ex. 2034, 6; Ex. 2051 ¶¶ 126–127). The bfloat16 format

"keep[s] the same 8 bit exponent as found in 32-bit IEEE floating point" but

"reduces precision significantly by using only a 7-bit mantissa, as opposed

to the 23-bit mantissa of IEEE 32-bit floating point." *Id.* at 57–58 (citing

Ex. 2041, 1–2). According to Patent Owner, Petitioner uses the TPUv2 and

TPUv3 products to "power all of its major products and services, including

IPR2021-00179
Patent 8,407,273 B2

Search, Translate, Photos, Assistant, and Gmail," saving Petitioner "at least $10 billion." *Id.* at 60–61 (citing Exs. 2011, 2015–18, 2040). Patent Owner argues that a nexus is presumed because the TPUv2 and TPUv3 products are coextensive with the challenged claims. *Id.* at 58, 61–66 (citing Pet. 19–20, 73; Ex. 1003 ¶¶ 231–232; Ex. 2011; Ex. 2016; Ex. 2041, 7; Ex. 2049; Ex. 2051 ¶¶ 129–136). Patent Owner further contends that, even if not presumed, a nexus is shown by "internal communications among [Petitioner's] own engineers and executives" and Petitioner's communications with Dr. Bates. *Id.* at 58–61 (citing Exs. 2007, 2011, 2016, 2035–39, 2041, 2042, 2047, 2048).

We have reviewed the cited evidence and do not find it probative of nonobviousness. First, we are not persuaded that a nexus should be presumed because the TPU products are coextensive with the Dockser/Tong-challenged claims, as Patent Owner contends. Petitioner and Mr. Goodin identify a number of other allegedly material features of the products that are not part of the claims: "Inter-Core Interconnect" (ICI), "Systolic array," "Two TensorCores per chip," "High-bandwidth memory," "Liquid cooling," and "Accelerated Linear Algebra compiler." *See* Reply 26; Ex. 1071 ¶¶ 86–92. We focus on the first two features in particular because the evidence of record characterizes them as significant.

Regarding the ICI of the TPU products, Dr. Patterson stated the following:

> So the *critical feature* of the supercomputer is the way they talk to each other – you know, the chips talk to each other. So what's called – [Petitioner] invented something called ICI. It's a custom Inter-Core Interconnect that they did custom for this design; it's not some standard that you buy off the shelf. It runs at about 500 gigabits per section in both directions per link, and there's four of those links.

46

IPR2021-00179
Patent 8,407,273 B2

Ex. 2012, 14:17–24 (emphasis added). An article authored by Dr. Patterson

and others[5] states that "[t]he *critical architecture feature* of a modern

supercomputer is how its chips communicate: what is the speed of a link;

what is the interconnect topology; does it have centralized versus distributed

switches; and so on," and the TPUv2 product has "four custom Inter-Core

Interconnect (ICI) links, each running at 496Gbits/s per direction."

Ex. 2016, 69; *see id.* at 70, 72, Fig. 2 (showing that ICI is part of the TPUv3

product as well); Ex. 1071 ¶ 87.

Regarding the systolic array architecture of the TPU products,

Dr. Patterson stated as follows:

> This systolic array idea brought back from the past that
> probably *cut the energy costs in half* for the multiplier, which is
> a lot of it. And then it was domain-specific. So a lot of the things
> that are complicated and area consuming on CPUs like branch
> predictors and caches – CPUs are covered with caches – they're
> not there because they're not helpful for this application. It was
> better to use more multiple units and dedicated memory units.
> So that *saves energy* and you could reuse the resources. So that
> was a *big success*.

Ex. 2012, 9:9–18 (emphasis added). He also agreed with an article

comparing the TPU to graphics processing units (GPUs) stating that "GPUs

incur high overhead in performance, area, and energy due to heavy

multithreading," whereas "[t]he systolic organization of TPUs captures the

data reuse while being simple by avoiding multithreading." *Id.* at

---

[5] We disagree with Patent Owner that the materials cited herein are merely
"marketing documents." *See* Conf. Tr. 13:18–24. Dr. Patterson's comments
are from an audio transcription of a YouTube video filed by Patent Owner
(Ex. 2012), and the article is from the publication *Communications of the
ACM* (Ex. 2016). *See* PO Resp. ix.

IPR2021-00179
Patent 8,407,273 B2

33:20–34:2. The article authored by Dr. Patterson and others includes similar statements. Ex. 2016, 70 ("The TPUv2 node of the supercomputer followed the main ideas of TPUv1: A large two-dimensional matrix multiply unit (MXU) *using a systolic array to reduce area and energy* plus large, software-controlled on-chip memories instead of caches." (emphasis added)), 74 (quoting the earlier article stating that the "systolic organization" of the TPU captures data reuse "while being simple by avoiding multithreading"); *see* Ex. 1071 ¶ 88.

Patent Owner contends that Petitioner fails to show that the "features are so critical that the TPUs are not 'essentially the invention.'" Sur-Reply 22. We are not persuaded. Petitioner does more than merely identify unclaimed features of the TPU products. Petitioner cites evidence, supported by the testimony of Mr. Goodin, showing that at least the ICI and systolic array architecture materially affect the TPU products' performance and were viewed as critical, successful features that were reasons for the TPU products' improved performance. *See Fox Factory*, 944 F.3d at 1374 ("[I]f the unclaimed features amount to nothing more than additional insignificant features, presuming nexus may nevertheless be appropriate."), 1375–76 (noting that the challenged patent and the patent owner's "marketing materials confirm that the forwardly protruding tooth tips, hook features, and mud clearing recesses each materially impacts the functioning of a chainring," such that "[f]or each of these features that the Board confirms is included in the [allegedly coextensive] chainrings, nexus can only be presumed between the . . . chainrings and a patent claim if the claim includes limitations relating to these features"). Patent Owner does not point to any evidence indicating that view to be factually incorrect. *See* Ex. 1072, 147:15–149:2 (Dr. Khatri opining that "the TPUs have the features that . . .

48

IPR2021-00179
Patent 8,407,273 B2

are in these claims" and stating that he has no opinion "about what the TPUs . . . might have in addition"). We find that a presumption of nexus is inappropriate with respect to the TPU products.

Regardless of whether a presumption of nexus applies, we are not persuaded that Patent Owner has made a sufficient showing of nexus for two reasons. First, Patent Owner's nexus arguments are premised on the TPUv2 and TPUv3 products' "low-precision" bfloat16 floating point format having an 8-bit exponent and 7-bit mantissa. PO Resp. 56–57 (arguing that Petitioner copied Dr. Bates's invention by implementing "execution units designed to perform operations using" the bfloat16 format), 61 (arguing commercial success due to the bfloat16 format using "low-precision multipliers"). But "[l]ow-precision arithmetic—including LPHDR multiplication—was not novel" and thus, cannot be the basis for a finding of nexus. *See* Reply 20; *Kao*, 639 F.3d at 1068; *Tokai Corp. v. Easton Enterprises, Inc.*, 632 F.3d 1358, 1369 (Fed. Cir. 2011) ("If commercial success is due to an element in the prior art, no nexus exists."). Tong taught such a format nine years before the '273 patent. Tong describes two different experiments emulating "different bitwidth FP units": one that "implements the IEEE-754 standard" and then varies the mantissa bitwidth (shown in Figure 6), and another that "implements the IEEE-754 standard" and then varies the exponent bitwidth (shown in Figure 7). Ex. 1008, 278–279. Thus, we agree with Mr. Goodin that a person of ordinary skill in the art "would have understood that Tong's Figure 6 . . . discloses a floating-point format with an 8-bit exponent (which is the exponent size of single-precision floating-point in the IEEE standard) . . . and a 7-bit mantissa." *See* Ex. 1071 ¶ 74 (citing Ex. 1008, 274, 278–279; Ex. 1003 ¶ 330). Tong discloses that the 8-bit exponent/7-bit mantissa format

IPR2021-00179
Patent 8,407,273 B2

produced useful results for both a "neural network trainer" application
(ALVINN) and a "speech recognition program" (Sphinx III). *See* Ex. 1008,
278–279; Ex. 1071 ¶ 75. Notably, these applications mirror Dr. Bates's
2013 presentation discussing ███████████████████████████████   *See*
Ex. 2005, 12–13.

　　　To be sure, the fact that the 8-bit exponent/7-bit mantissa format was
known in the prior art is not dispositive. *See WBIP*, 829 F.3d at
1330–31. "Commercial success, for example, may be linked to an
individual element or, in other circumstances, it could be linked to the
inventive combination of known elements." *Id.* at 1332. Patent Owner,
however, has not identified and explained a specific combination of
elements (as opposed to just the "low-precision" bfloat16 floating point
format) that could serve as a nexus for the evidence of alleged copying and
commercial success.[6] *See* PO Resp. 56–57, 61.

　　　Second, when arguing that the TPUv2 and TPUv3 products meet the
X and Y percentage limitations of claim 1, Patent Owner points to the
bfloat16 floating point format with a 7-bit mantissa and states that Petitioner
argues in the Petition (with respect to Dockser) that "utilizing 7 bits of
mantissa in multiplication operations results in a minimum of 12% of valid
floating point 32 inputs producing at least 0.39% relative error compared to
the exact mathematical calculation of a full-precision multiplication on those
same inputs." *Id.* at 65–66 (citing Pet. 73; Ex. 2051 ¶¶ 135–136). Patent
Owner does not explain, however, how the cited analysis of how Dockser

---

[6] To the extent Patent Owner alleges commercial success from Petitioner
"[u]sing hundreds of thousands of these low-precision TPU v2 and v3
devices," we note again that claim 1 recites "at least one" LPHDR execution
unit and thus is satisfied by a single unit. *See* PO Resp. 61.

IPR2021-00179
Patent 8,407,273 B2

teaches the X and Y percentage limitations applies to the TPUv2 and TPUv3 products. As Petitioner points out, Patent Owner "does not even attempt to demonstrate that [Petitioner's] TPUs perform the same bit-dropping techniques as Dockser." *See* Reply 27. Patent Owner has not established a sufficient nexus between the alleged copying and commercial success of the TPUv2 and TPUv3 products and the Dockser/Tong-challenged claims.

Finally, we note that Patent Owner's evidence of commercial success would be unpersuasive even if Patent Owner had established a nexus. Patent Owner's sole support for the alleged commercial success of the TPUv2 and TPUv3 products is an article in *PCMag India*. *See* PO Resp. 61 (citing Ex. 2017). That article states in relevant part:

> As AI and Machine Learning advanced, [Petitioner] was faced with the proposition of doubling the number of data centers it operates to handle the increased workloads. Building multiple data centers to handle the growing amounts of data sets *could have easily cost the company anything above $10 billion*.
>
> So [Petitioner] decided to take another route by optimizing its current data centers for better efficiency. The search giant in late 2013 decided to produce a custom chip in order to improve cost-performance over the GPUs it had deployed. Thus *the first TPU* was born.

Ex. 2017, 3 (emphasis added). The above statement as to how much "could have" been saved is too speculative to reach a conclusion of commercial success. *See id.* It also pertains to Petitioner's "first" TPU, not the TPUv2 and TPUv3 products that use the bfloat16 format. *See id.*; Ex. 2016, 69 ("Quantized arithmetic—8-bit integer instead of 32-bit floating point (FP)— can work for inference like in TPUv1 but reduced-precision training is an active research area.").

IPR2021-00179
Patent 8,407,273 B2

Accordingly, we find that Patent Owner's evidence of alleged copying and commercial success does not weigh in favor of nonobviousness of the Dockser/Tong-challenged claims.

f) Conclusion

Based on all of the evidence of record, including evidence of objective indicia of nonobviousness submitted by Patent Owner, we determine that claim 1 would have been obvious based on Dockser under 35 U.S.C. § 103(a).

3. Claims 2, 21–24, 26, and 28

Claims 2, 21–24, 26, and 28 depend from claim 1. Claim 2 recites that "the at least one first LPHDR execution unit comprises at least part of an FPGA." Dockser discloses an FPP that is part of "a field programmable gate array (FPGA) or other programmable logic component." Ex. 1007 ¶ 35; *see* Pet. 36.

Claim 21 recites that "the device includes memory locally accessible to the at least one first LPHDR execution unit." Petitioner argues that local registers 110 are locally accessible to the FPP. Pet. 36–37 (citing Ex. 1007 ¶¶ 15–16).

Claim 22 recites that "the device is implemented on a silicon chip" and claim 23 recites that "the device is implemented on a silicon chip using digital technology." Petitioner cites Dockser's disclosure that the FPP can be part of an ASIC and "process[es] . . . bits." *Id.* at 37 (citing Ex. 1007, code (57), ¶ 35). Petitioner further argues that "[i]mplementing Dockser's integrated circuit on a silicon chip would have been the typical, conventional, and obvious implementation." *Id.* (citing Ex. 1003 ¶ 316;

52

IPR2021-00179
Patent 8,407,273 B2

Ex. 1054 ¶ 5 ("[a]n integrated circuit conventionally comes in the form of a silicon chip"); Ex. 1055 ¶ 5 ("Integrated circuits (chips) are generally made of silicon on which electronic circuits are fabricated.")).

Claim 24 recites that "the device further comprises a digital processor adapted to control the operation of the at least one first LPHDR execution unit." Petitioner argues that "Dockser's computing system . . . includes a 'main processor' which a [person of ordinary skill in the art] would have understood is a 'digital processor' . . . because it communicates via 'bits.'" *Id.* (citing Ex. 1007 ¶¶ 15, 25, 35). Petitioner further argues that even if not, using a digital processor "would have been the typical and obvious implementation of Dockser's 'conventional processor.'" *Id.* at 37–38 (citing Ex. 1007 ¶ 35; Ex. 1023, col. 1, ll. 16–48). According to Petitioner, "Dockser's main processor is 'adapted to control the operation of'" the FPP "by specifying its precision level by writing 'subprecision select bits' to the FPP's '*control* register.'" *Id.* at 38 (citing Ex. 1007 ¶¶ 18, 25).

Claim 26 recites that "the device is part of a mobile device." Petitioner cites Dockser's disclosure that the computing system can be part of a wireless telephone, personal digital assistant (PDA), pager, or other type of device. *Id.* (citing Ex. 1007 ¶ 3).

Claim 28 recites that "the at least one first LPHDR execution unit represents numbers using a floating point representation." Dockser teaches an FPP operating on numbers using a floating point representation (with an exponent and mantissa). Ex. 1007 ¶¶ 1–2, 15–22; *see* Pet. 9–13, 38.

Patent Owner does not argue separately dependent claims 2, 21–24, 26, and 28 in its Response, only disputing Petitioner's contentions with respect to parent claim 1. PO Resp. 20–24. We have reviewed Petitioner's contentions regarding claims 2, 21–24, 26, and 28, which are consistent with

IPR2021-00179
Patent 8,407,273 B2

the disclosure of Dockser and supported by the testimony of Mr. Goodin,

taken into account evidence of objective indicia of nonobviousness

submitted by Patent Owner, and are persuaded that Petitioner has proven, by

a preponderance of the evidence, that dependent claims 2, 21–24, 26, and 28

would have been obvious based on Dockser under 35 U.S.C. § 103(a). *See*

Pet. 36–38; Ex. 1003 ¶¶ 311–322; *supra* Section II.D.2(e).

E. Obviousness Ground Based on Dockser and Tong (Claims 1, 2, 21–24, 26, 28, 32, and 33)

1. Tong

Tong is an IEEE journal article entitled "Reducing Power by

Optimizing the Necessary Precision/Range of Floating-Point Arithmetic."

Ex. 1008, 273. Tong teaches reducing power consumption by minimizing

the bitwidth representation of floating-point data. *Id.* According to Tong,

using a variable bitwidth floating-point unit saves power. *Id.*

2. Claims 1, 2, 21–24, 26, and 28

Petitioner argues that the subject matter of claim 1 would have been

obvious over the combination of Dockser and Tong[7] for reasons similar to

those analyzed with respect to Dockser alone. *See* Pet. 39–43. Additionally,

Petitioner asserts that "Tong, like Dockser, . . . confirms that the number of

mantissa bits used in a high-dynamic-range floating-point execution unit was

a well-known result-effective variable impacting power consumption and

---

[7] Petitioner provides evidence supporting its contention that Tong is a prior
art printed publication under 35 U.S.C. § 102(b). *See* Pet. 38–39 (citing
Ex. 1025 ¶¶ 8–11; Ex. 1026, 27; Ex. 1027, 27). Patent Owner does not
assert otherwise in its Response, and we agree that Tong is prior art for the
reasons stated by Petitioner.

IPR2021-00179
Patent 8,407,273 B2

precision." *Id.* at 39–40 (citing Ex. 1008, 273–278; Ex. 1003 ¶¶ 324–325).

Petitioner relies on Tong's Figure 6, reproduced below with Petitioner's

annotations. *Id.* at 40.
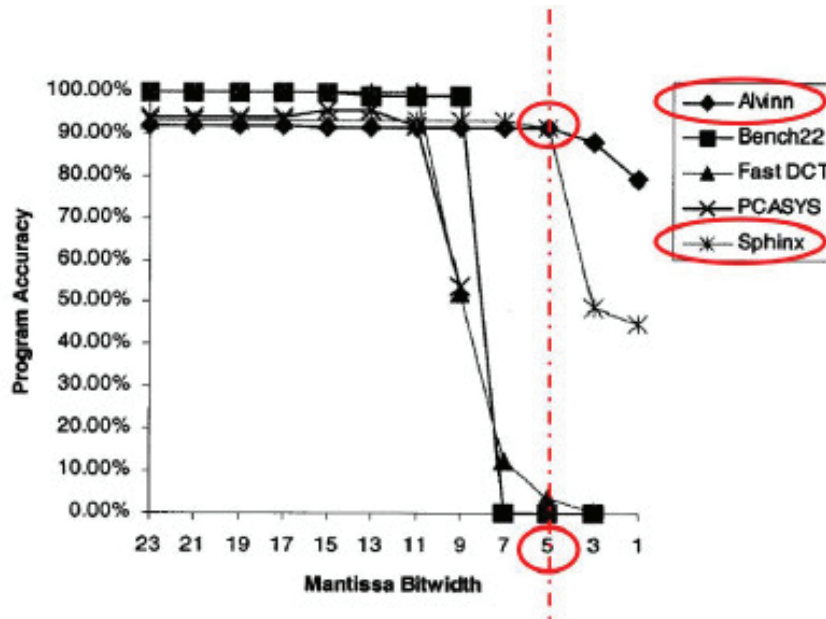
Tong's Figure 6, above, is a line graph showing program accuracy, from 0%

to 100%, on the vertical axis and mantissa bitwidth, from 1 to 23, on the

horizontal axis. *Id.* at 39–40. The figure shows this data for five programs:

ALVINN, Bench22, Fast DCT, PCASYS, and Sphinx III. *Id.* The programs

implement different signal-processing tasks. *See, e.g.*, Ex. 1008, 278

(Table IV). ALVINN, for example, is a neural network trainer that uses

backpropagation. *Id.* And Sphinx III is a speech-recognition program. *Id.*

      Petitioner argues that "Tong teaching that [a precision level] retaining

5 mantissa fraction bits is sufficient in some applications (including

ALVINN and Sphinx [III]) would have motivated a [person of ordinary skill

in the art] to configure Dockser's FPP . . . to operate at a selected precision

level retaining as few as 5 mantissa fraction bits." Pet. 41–42. According to

Petitioner, one of ordinary skill in the art would have done so "to conserve

power when running those applications, or others empirically determined to

IPR2021-00179
Patent 8,407,273 B2

not require greater precision using Tong's techniques." *Id.* (citing Ex. 1003 ¶ 334). Petitioner also concludes that "[d]etermining the optimum range of imprecision to achieve the best power reduction without sacrificing accuracy for a particular application was a matter of routine optimization of a result-effective variable." *Id.* at 43.

Patent Owner argues that, because Dockser lacks an LPHDR execution unit and Tong does not remedy that deficiency, the Dockser-Tong combination does not render obvious any challenged claim. PO Resp. 24; Sur-Reply 9. We disagree for the reasons discussed above regarding Petitioner's challenge based on Dockser alone. *See supra* Section II.D.2.

Patent Owner further argues that Tong does not teach using only 5 mantissa bits, which means that the Dockser-Tong combination lacks the imprecision required to meet the X and Y values recited in limitation *1B2*. PO Resp. 24. In Patent Owner's view, "Tong teaches that at least 11 mantissa bits are required for consistent performance, even for 'programs dealing with human interfaces [that] process sensory data with intrinsically low resolutions.'" *Id.* (quoting Ex. 1008, 278). According to Patent Owner, "Petitioner makes no effort to show that arithmetic with the 'required' 11-bit mantissa is an LPHDR operation and meets the X/Y limitations of the claims." *Id.* at 25.

Yet Patent Owner's argument about an 11-bit mantissa (*id.* at 24–25) does not squarely address Petitioner's obviousness rationale—i.e., that one of ordinary skill in the art would have been motivated to configure Dockser's FPP to operate at a "selected precision level" according to Tong (Pet. 41–43). Specifically, Petitioner's rationale proposes using a selected precision level "retaining as few as 5 mantissa fraction bits." *Id.* at 41–42. Thus, we disagree with Patent Owner's argument about an 11-bit mantissa.

IPR2021-00179
Patent 8,407,273 B2

See PO Resp. 24–25; Sur-Reply 9. For this reason, we do not credit Dr. Khatri's testimony on this point. Ex. 2051 ¶ 77.

Petitioner used a software program to determine the relative error when retaining 5 mantissa fraction bits—i.e., the number of bits that Dockser would use under Petitioner's proposed combination with Tong. Pet. 42 (citing Ex. 1003 ¶ 336). Also, the algebraic analysis cited by Petitioner uses 5 mantissa fraction bits. *Id.* (citing Ex. 1003 ¶ 336).

Petitioner adequately shows that the X and Y limitations are met for the 5 mantissa bit rationale. *Id.* at 41–43. In particular, Mr. Goodin's software program shows that "when retaining 5 mantissa fraction bits in view of Tong, Dockser's register bit-dropping technique produces at least Y=0.05% relative error for 99.45% of possible valid inputs, and Dockser's logic bit-dropping technique produces at least Y=0.05% relative error for 99.47% of possible valid inputs," and the algebraic analysis shows that "Dockser's register bit-dropping technique when retaining 5 mantissa fraction bits produces a minimum of 1.56% relative error (greater than Y=0.05%) for over 12% of possible valid inputs (greater than X=5%)." Ex. 1003 ¶ 336. After review of the arguments and evidence on these issues, we credit Mr. Goodin's algebraic analysis and software-based analysis. *See id.*, App'x I.B–D.

Although Patent Owner does not present argument specifically directed to Mr. Goodin's software-based or algebraic analysis, Patent Owner argues that one of ordinary skill in the art would not have chosen a 5-bit mantissa: "Tong's experimental results show that using fewer than 11 mantissa bits unacceptably reduces accuracy for the benchmark applications *that were specifically selected* (from a category of programs) for their high tolerance for imprecision." PO Resp. 25 (citing Ex. 2051 ¶ 78). To

IPR2021-00179
Patent 8,407,273 B2

illustrate this point, Patent Owner annotates Tong's Figure 6, reproduced below.
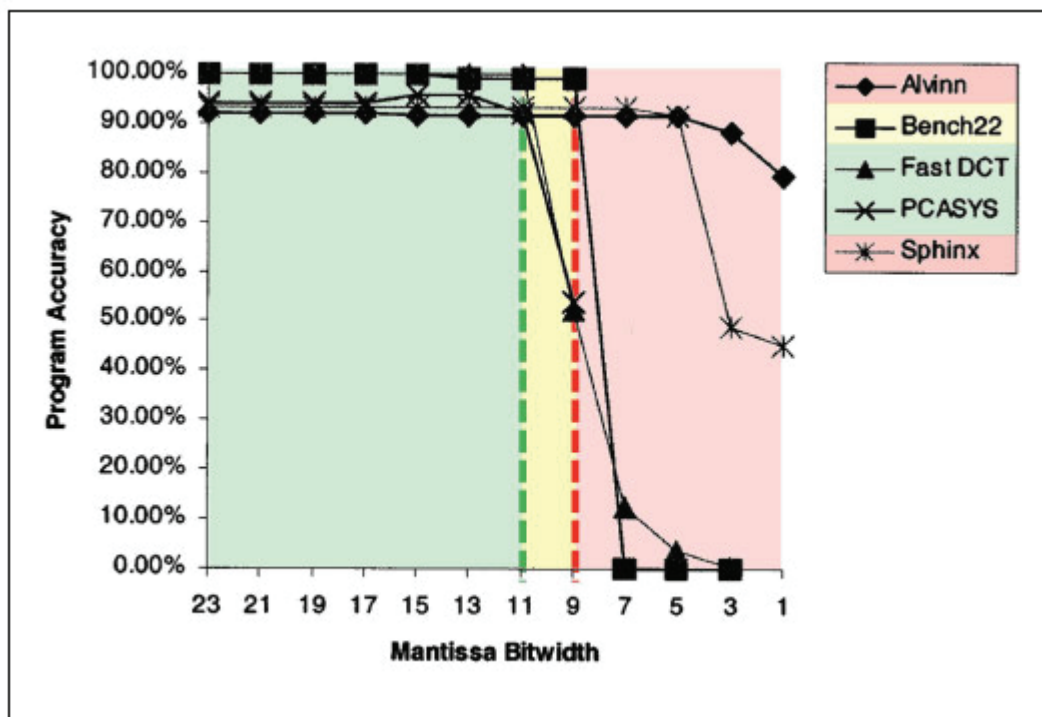
Tong's Figure 6, above, shows the experimental results with color annotations indicating "the precision levels suitable for general-purpose operation (green), those at which a significant percentage of applications begin to produce unacceptable results (yellow), and finally, levels of precision that are unsuitable for most testing benchmarks (red)." *Id.* at 25–26 (citing Ex. 2051 ¶ 79).

Patent Owner argues that there is no motivation to combine Dockser and Tong because "there is no evidence to show that a [person of ordinary skill in the art] would have chosen Tong's 5-bit examples over the 7, 9, or 11-bit examples." *Id.* at 26; *see also* Sur-Reply 9–10. According to Patent Owner, "Tong does not disclose any extra capability that would have motivated a [person of ordinary skill in the art] to use it with Dockser." PO Resp. 26 (citing Ex. 2051 ¶¶ 81–82). In Patent Owner's view, Petitioner

58

IPR2021-00179
Patent 8,407,273 B2

relies on hindsight to choose 5 bits and does not offer any differences between Dockser and the claimed invention or other non-hindsight motivation to combine with Tong. *Id.* According to Patent Owner, "Tong teaches that 'it appears inevitable that some fraction of our operands will require full IEEE standard precision,'" which means that a person of ordinary skill in the art would not have been motivated to operate only in 5-bit, 7-bit, or 9-bit modes as Petitioner suggests. Sur-Reply 10 (quoting Ex. 1008, 280).

We disagree with Patent Owner's arguments directed to the rationale for using 5 mantissa bits. *See* PO Resp. 24–26; Sur-Reply 9–10. Petitioner's reasoning for the 5-bit example is "to conserve power when running [certain] applications, or others empirically determined to not require greater precision using Tong's techniques." Pet. 41–42 (citing Ex. 1003 ¶ 334). For example, in the ALVINN and Sphinx III line plots in Tong's Figure 6, Petitioner added a dashed red line extending vertically through the data points with a mantissa bitwidth of 5. *Id.* at 40. Petitioner asserts that, for these programs, Tong teaches that "the accuracy does not change significantly with as few as 5 mantissa [fraction] bits." *Id.* (quoting Ex. 1008, 278). In Petitioner's view, Tong omits unnecessary bits to reduce waste and power consumption. *Id.* at 40–41. Petitioner characterizes Tong as "[h]aving empirically determined the minimum number of mantissa bits necessary to maintain acceptable accuracy of particular applications." *Id.* at 41 (citing Ex. 1008, 273, 274, 279, 284).

Petitioner's assertions are adequately supported by the record. We are persuaded by Petitioner's reliance on Tong's Figure 6 to determine how many bits are needed in a particular application. *See id.* at 40–42. That is, Petitioner's choice is not arbitrary or solely based on the '273 patent's

59

IPR2021-00179
Patent 8,407,273 B2

written description alone. *See* PO Resp. 24–26. Rather, Petitioner's choice

of 5 bits is based on Tong's experimental results for *certain* applications.

See Pet. 40–42. We emphasize *certain* because Petitioner is not saying that

5-bits is suitable for every application. *See id.* Full precision may be needed

some of the time. *See* Ex. 1008, 280; Sur-Reply 9–10. But, even if Tong

teaches that some applications begin to produce unacceptable results below

11-bits, as Patent Owner notes (PO Resp. 24–25), Tong teaches that 5-bits

would be enough for certain ones: ALVINN and Sphinx III shown in Tong's

Figure 6. Pet. 40–42.

Patent Owner argues that "Tong lists those two applications amongst

a great many other disclosed applications that all require full precision, and

that a [person of ordinary skill in the art] would therefore not be motivated

to configure Dockser to operate with a 5-bit mantissa just to support those

two applications given the far larger number of disclosed applications that

require full precision." Sur-Reply 9; PO Resp. 24–26.

We disagree with Patent Owner's argument here because the record

adequately supports Petitioner's rationale that "[d]etermining the optimum

range of imprecision to achieve the best power reduction without sacrificing

accuracy for a particular application was a matter of routine optimization of

a result-effective variable," and Tong, in particular, shows an optimization

for a particular program. Pet. 41–43. The record adequately supports

Petitioner's reasoning here: Dr. Khatri testified that "exact computing . . .

takes up a lot of power." Ex. 1072, 47:2–5. Mr. Goodin explained that

reducing precision may reduce power consumption. *See, e.g.*, Ex. 2043,

34:21–35:4; *see also id.* at 30:22–31:9 (discussing how selectable precision

may achieve power savings). And we credit Mr. Goodin's explanation of

how Tong achieves this for certain applications. *See, e.g.*, Ex. 1003

60

IPR2021-00179
Patent 8,407,273 B2

¶¶ 31–33 (citing Ex. 1008, 273, 277–279), 334, 337 (citing Ex. 1008, 278).
That is, Petitioner's proposed combination would omit unnecessary bits to reduce waste and power consumption. Pet. 41–43.

Thus, we find Patent Owner's arguments regarding the combination with Tong unavailing. *See* PO Resp. 24–26; Sur-Reply 9–10. For the same reasons, we do not credit Dr. Khatri's testimony on this point. Ex. 2051 ¶¶ 77–82.

For the reasons discussed, we determine that Petitioner provides articulated reasoning, supported by rational underpinnings, why one of ordinary skill in the art would have combined Dockser and Tong. *See KSR*, 550 U.S. at 418. Based on the Petition, we find that a person of ordinary skill in the art would have modified Dockser with Tong to arrive at the device recited in claim 1 with a reasonable expectation of success. Based on all of the evidence of record, including evidence of objective indicia of nonobviousness submitted by Patent Owner, we determine that claim 1 would have been obvious based on Dockser and Tong under 35 U.S.C. § 103(a).

Petitioner asserts that the subject matter recited in claims 2, 21–24, 26, and 28 would have been obvious over the combination of Dockser and Tong for reasons similar to those presented in the ground based on Dockser alone. *See* Pet. 41–43. For the reasons discussed above and the reasons discussed for Dockser alone, Petitioner's contentions are supported by Mr. Goodin's testimony and are persuasive. *See id.*; *supra* Section II.D.3. Patent Owner does not present arguments specifically directed to these claims. *See* PO Resp.; Sur-Reply. We disagree with Patent Owner's arguments regarding claim 1 based on the full trial record for the same reasons explained above. *See supra* Section II.D.2. Based on all of the evidence of

IPR2021-00179
Patent 8,407,273 B2

record, including evidence of objective indicia of nonobviousness submitted

by Patent Owner, we determine that claims 2, 21–24, 26, and 28 would have

been obvious based on Dockser and Tong under 35 U.S.C. § 103(a).

3. Claim 32

Claim 32 depends from claim 1 and recites that "the device is adapted

to perform nearest neighbor search." Petitioner argues that the combined

device in the asserted combination "implements Dockser's FPP using the

precision Tong teaches for neural-network image-classification applications

including ALVINN," and a person of ordinary skill in the art would have

been motivated to perform "nearest neighbor search" as claimed as well

because that type of search "was commonly used and well known to be

advantageous for image-classification (including neural-network)

applications." Pet. 43–44 (citing Ex. 1003 ¶ 340; Ex. 1056 ¶¶ 1–4;

Ex. 1057, col. 1, ll. 15–50). Patent Owner does not argue separately

dependent claim 32 in its Response, only disputing Petitioner's contentions

regarding parent claim 1 and the alleged motivation to combine Dockser and

Tong, which we address above. PO Resp. 24–26; *see supra* Section II.E.2.

Based on all of the evidence of record, including evidence of objective

indicia of nonobviousness submitted by Patent Owner, we determine that

claim 32 would have been obvious based on Dockser and Tong under

35 U.S.C. § 103(a). *See* Pet. 43–44; Ex. 1003 ¶ 340.

4. Claim 33

Claim 33 recites a "device comprising a computer processor and a

computer-readable memory storing computer program instructions, wherein

the computer program instructions are executable by the processor to

IPR2021-00179
Patent 8,407,273 B2

emulate a second device comprising" an LPHDR execution unit having the same properties as recited in claim 1. Petitioner asserts that Tong teaches emulating, in software, different bitwidth FP units to determine application accuracy. Pet. 44 (citing Ex. 1008, 278). According to Petitioner, based on this teaching, one of ordinary skill in the art would have been motivated "to emulate the Dockser/Tong device . . . in software to assess the accuracy of applications running on the device at selected precision levels, and a [person of ordinary skill in the art] would have had a reasonable expectation of success in doing so." *Id.* (citing Ex. 1003 ¶¶ 341–342).

Petitioner asserts that "[w]hen the Dockser/Tong device is 'emulated in software' per Tong's teachings, . . . the conventional and obvious implementation of such 'software' is 'computer program instructions' stored in 'a computer-readable memory' and 'executable by [a] processor' within a 'device' (*e.g.*, a computer) as claimed." *Id.* (citing Ex. 1003 ¶ 343; Ex. 1007 ¶ 36; Ex. 1029 ¶¶ 8, 9, 14; Ex. 1001, col. 26, ll. 2–4). For reasons similar to those with respect to claim 1, Petitioner argues that a person of ordinary skill in the art would have been motivated by Tong to use a computer to emulate the combined teachings of Dockser and Tong, "meeting the 'second device' [recited in claim 33]), which comprises Dockser's execution unit." *Id.* at 44–45 (citing Ex. 1003 ¶ 344). Petitioner argues that Tong emulates a device in the same way as the '273 patent does:

> Like the '273 patent's only examples of "emulat[ing] a . . . device" comprising LPHDR execution unit(s), Tong emulates the device by executing a software application program that runs thereon with arithmetic operations in the application program replaced by software simulations of the reduced-precision arithmetic the device would perform.

63

IPR2021-00179
Patent 8,407,273 B2

Id. at 45 (citing Ex. 1001, col. 17, ll. 3–17, col. 18, l. 52–col. 19, l. 12;
Ex. 1008, 278; Ex. 1003 ¶¶ 345–346).

Patent Owner does not argue separately claim 33 in its Response, only
disputing Petitioner's contentions regarding claim 1 and the alleged
motivation to combine Dockser and Tong, which we address above.
PO Resp. 24–26; *see supra* Section II.E.2. After review of the arguments
and evidence on these issues, we agree with Petitioner and credit
Mr. Goodin's testimony. *See* Ex. 1003 ¶¶ 341–348.

In particular, the '273 patent describes code written in the
C programming language to perform the same arithmetic operations in the
same order using the same method as one implemented in hardware.
Ex. 1001, col. 18, l. 52–col. 19, l. 12, *cited in* Pet. 45. We agree with
Petitioner that Tong similarly uses a computer program to simulate reduced
precision arithmetic that a hardware unit would perform. *See* Pet. 45. For
example, Tong uses an "FP software emulation package" to simulate
reduced precision arithmetic and determine "the relationship between
program accuracy and number of bits in FP representation." Ex. 1008, 278,
cited in Pet. 45.

Petitioner's obviousness rationale—i.e., "to assess the accuracy of
applications running on the device at selected precision levels"—is
adequately supported by the record: Tong explains that the emulation
package was used to assess "different mantissa and exponent bitwidths."
See Pet. 44 (citing Ex. 1008, 278). Petitioner's assertion that software is
computer program instructions stored in a computer-readable memory and
executable by a processor within a computing device is adequately
supported by at least Mr. Goodin's testimony and the evidence that he cites.

IPR2021-00179
Patent 8,407,273 B2

Id. at 44–45 (citing Ex. 1003 ¶ 343; Ex. 1007 ¶ 36; Ex. 1029 ¶¶ 8, 9, 14;

Ex. 1001, col. 26, ll. 2–4).

For the same reasons discussed with respect to claim 1, Petitioner has

shown that the subject matter recited in claim 33 that is also recited in

claim 1 would have been obvious over the Dockser-Tong combination. *See*

supra Section II.E.2.

We determine that Petitioner provides articulated reasoning, supported

by rational underpinnings, why one of ordinary skill in the art would have

combined Dockser and Tong. *See KSR*, 550 U.S. at 418. Based on the

Petition, we find that a person of ordinary skill in the art would have

modified Dockser with Tong to arrive at the device recited in claim 33 with

a reasonable expectation of success. Based on all of the evidence of record,

including evidence of objective indicia of nonobviousness submitted by

Patent Owner, we determine that claim 33 would have been obvious based

on Dockser and Tong under 35 U.S.C. § 103(a).

F.   Obviousness Ground Based on Dockser and MacMillan (Claims 1–26, 28, 36–61, and 63)

1.   MacMillan

MacMillan is entitled "Circuit for Enhancing Performance of a

Computer for Personal Use." Ex. 1009, code (54). MacMillan discloses

using Single Instruction Multiple Data (SIMD) parallel-processing

architectures for "adding supercomputer performance to a computer for

personal use." *Id.* at col. 5, ll. 22–54. MacMillan's computer system

comprises a "Host CPU" (i.e., "a 386, 486 or Pentium[] processor") and

SIMD-random access memory (SIMD-RAM) devices. *Id.* at col. 9,

ll. 11–19, 30–31, Figs. 2, 3.

65

IPR2021-00179
Patent 8,407,273 B2

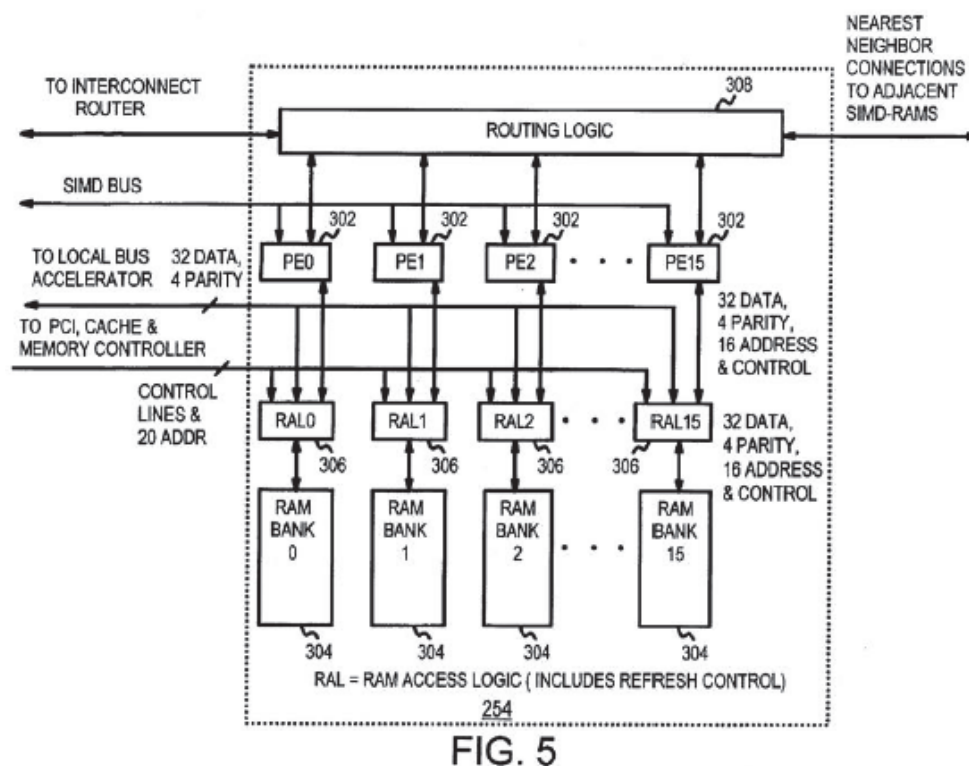Figure 5 of MacMillan is reproduced below.

FIG. 5

Figure 5 depicts SIMD-RAM device 254 comprising a plurality of
DRAMs 304 coupled to a plurality of processing elements 302 (PE0, PE1,
PE2, etc.) connected to SIMD bus 240. *Id.* at col. 12, ll. 35–43. MacMillan
discloses that "[e]ach PE contains a 32-bit wide data path and can perform
atomic operations on bits, bytes, 16-bit words, and 32-bit words," "each
DRAM 304 is independently addressed by its PE," "[i]nteger and floating
point accelerators could be included in each PE," and "[e]xecution autonomy
is provided, in which specific PEs can be excluded from executing specific
instructions." *Id.* at col. 12, ll. 47–59. MacMillan describes an example
architecture of a system with 256 PEs, but states that the disclosed
architecture "allows scaling to higher or lower density chips with more or
fewer PEs." *Id.* at col. 12, l. 60–col. 13, l. 4, col. 13, ll. 38–41, col. 16,
ll. 20–22.

66

IPR2021-00179
Patent 8,407,273 B2

2. Claim 8

Petitioner contends that claims 1–26, 28, 36–61, and 63 are unpatentable over Dockser and MacMillan under 35 U.S.C. § 103(a). Pet. 45–56. We begin with claim 8, which is illustrative of the disputed issues addressed by the parties in their papers.

Claim 8 depends from claim 1 and recites that "the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide" (the "exceeds" limitation). Petitioner relies on MacMillan for its teachings regarding multiple floating-point execution units, and argues that a person of ordinary skill in the art "would have been motivated to use Dockser's FPP to implement each 'floating-point accelerator' in the parallel PEs of MacMillan's SIMD architecture to increase performance speed as MacMillan teaches while lowering power consumption as Dockser teaches." *Id.* at 5, 47–48 (citing Ex. 1003 ¶¶ 354–355). According to Petitioner, "using Dockser's FPP as the floating-point accelerator in MacMillan's PE would have achieved the predictable result of enabling the PEs to perform reduced-precision floating-point arithmetic as taught by Dockser at reduced power." *Id.* at 48 (citing Ex. 1003 ¶ 356; Ex. 1009, col. 12, ll. 47–49; Ex. 1024, col. 1, ll. 43–58). Petitioner further contends that MacMillan's SIMD parallel processing is used in "the same types of computers for which Dockser's FPP provides '[p]ower management'" and would be beneficial for the same types of "'graphics applications' for which Dockser teaches its FPP can beneficially save power by reducing unnecessary precision." *Id.* at 47–48

IPR2021-00179
Patent 8,407,273 B2

(citing Ex. 1003 ¶¶ 353, 357; Ex. 1007 ¶ 3; Ex. 1009, col. 1, ll. 6–9, col. 5, ll. 22–45, col. 7, ll. 14–34).

With respect to the "exceeds" limitation in particular, Petitioner asserts that MacMillan teaches 256 PEs, and the Dockser-MacMillan combination would have a "single Host CPU floating-point unit" and at least one FPP in each of the PEs, of which there can be "256 or more, up to 'tens of thousands.'" *Id.* at 49–51 (quoting Ex. 1009, col. 2, ll. 13–15; citing Ex. 1003 ¶¶ 361, 369, 370, 374). Thus, in Petitioner's proposed combined device based on Dockser and MacMillan, the number of LPHDR execution units (Dockser FPPs) "exceeds its number (one) of traditional-precision execution units (the single Host CPU floating-point unit) by . . . more than 100." *Id.* at 50–51 (citing Ex. 1003 ¶¶ 369, 370, 374).

Patent Owner responds that a person of ordinary skill in the art would not have been motivated to combine the teachings of Dockser and MacMillan in the manner asserted by Petitioner. PO Resp. 34–36. Patent Owner argues that "Dockser is focused on the objective of reducing power consumption," whereas "MacMillan is focused on a parallel architecture that increases computational power, and not focused on reducing power consumption." *Id.* at 34–35 (citing Ex. 1009, col. 3, ll. 4–6; Ex. 2051 ¶¶ 101–102). Patent Owner further contends that "[i]ncorporating Dockser's FPPs into MacMillan would . . . defeat MacMillan's stated objective of achieving a high-scale SIMD computer architecture at 'lower system cost.'" *Id.* at 35 (citing Ex. 1009, col. 5, ll. 58–59; Ex. 2051 ¶ 103). According to Patent Owner, "Dockser's FPPs are even *larger* than traditional full-precision execution units because of the control circuitry needed to implement the selectable subprecision modes," such that "replacing the full-precision execution units of MacMillan with Dockser FPP units would

68

IPR2021-00179
Patent 8,407,273 B2

require additional circuitry and chip space and would therefore *increase* costs, while providing no benefit." *Id.* at 34–35 (citing Ex. 2051 ¶ 104); *see also* Sur-Reply 14–15.

Petitioner argues in its Reply that the asserted combination would achieve both benefits of increasing performance speed (based on MacMillan) and lowering power consumption (based on Dockser), and that Patent Owner fails to provide supporting evidence for its argument that incorporating Dockser's FPP into the MacMillan device would increase costs, and even if it had, economic reasons are not indicative of nonobviousness. Reply 9–10 (citing Ex. 1071 ¶¶ 30–31).

After reviewing the parties' arguments and evidence, and the full disclosures of Dockser and MacMillan, we agree with Patent Owner. Dockser teaches a single FPP with "selectable subprecision" that is capable of both full precision and reduced precision operations. *See* Ex. 1007, code (57), ¶ 15. The FPP includes additional control circuitry, such as floating-point controller 130 and control register 137, that allows selection of a subprecision for floating-point operations. *Id.* ¶¶ 15, 18, Fig. 1. Dockser describes "a variety of ways" in which "subprecision select bits may be used to reduce the precision of a floating-point operation." *Id.* ¶¶ 18, 23–34. Reduced precision can be achieved, with resultant power savings, either by removing power from "floating-point register elements for the excess bits of the fraction that are not required to meet the precision specified by the subprecision select bits" or "removing power to the logic in the floating-point operator 140 that remains unused as a result of the subprecision selected." *Id.* ¶¶ 4–7, 26–27. Mr. Goodin agrees that Dockser's FPP requires additional control circuitry to implement its selectable subprecision. Ex. 2043, 43:21–45:9.

IPR2021-00179
Patent 8,407,273 B2

MacMillan, by contrast, is heavily focused on avoiding increased costs associated with added complexity. MacMillan states that "[i]f supercomputing performance could be achieved in low cost computers, such as personal computers, this could dramatically expand the market for personal computers." Ex. 1009, col. 1, ll. 21–24. SIMD computing capability can be added to personal computers, "resulting in much higher performance computing at moderate cost." *Id.* at col. 8, ll. 51–55. MacMillan, however, describes numerous limitations of known SIMD designs at the time, many of which relate to increased costs from added complexity (e.g., additional registers), size, etc. *See, e.g., id.* at col. 2, ll. 56–61 ("hav[ing] a large register set on each PE . . . increases die area per PE, resulting in higher costs per PE"), col. 2, ll. 62–65 ("Adding pins can reduce the PE-to-memory bottleneck, but leads to increased packaging costs. It may also require tighter geometries or increased numbers of layers on printed circuits boards, further increasing costs."), col. 3, ll. 2–4 ("Adding output buffers to drive increased pin counts also increases power dissipation and hence power supply capacity and cost."), col. 4, ll. 39–40 ("The PE array[] disk may be limited in size due to technology or cost factors.").

Importantly, MacMillan states:

> Overcoming the above limitations would allow supercomputing performance to be provided in a *low cost* computer system for personal use, dramatically expanding the potential market for systems with supercomputer performance. To meet the cost objectives, the SIMD capabilities *should not add significant complexity to the architecture of a computer system for personal use.* The present invention addresses the above needs.

Id. at col. 5, ll. 38–45 (emphasis added). MacMillan's disclosed architecture—with multiple, less costly PEs operating in parallel to increase computational power—is expressly intended to account for these limitations.

70

IPR2021-00179
Patent 8,407,273 B2

See id. at col. 5, ll. 44–45, col. 5, ll. 58–61 (the disclosed system with "shared memory results in lower system cost and more flexibility in using memory"), col. 6, ll. 1–10 ("Another aspect of the invention is that it shows how parallel SIMD processing can be added inexpensively to existing system architectures. This cost-effective addition of parallel processing can be made to popular, low cost computer systems running popular operating systems . . . ."), col. 6, ll. 41–45, col. 16, ll. 14–27.

In determining how a person of ordinary skill in the art would have understood those disclosures, we credit the testimony of Dr. Khatri, which is consistent with, and a logical reading of, Dockser and MacMillan. *See Icon Health & Fitness, Inc. v. Strava, Inc.*, 849 F.3d 1034, 1041 (Fed. Cir. 2017) (stating that the Board "is permitted to weigh expert testimony and other record evidence and, in so doing, rely on certain portions of an expert's declaration while disregarding others"). Specifically, the circuitry implementing Dockser's selectable subprecision is beyond that of a typical full precision execution unit. Dr. Khatri testifies that

> [b]ecause Dockser's execution unit is capable of full-precision, and uses conventional arithmetic units, a [person of ordinary skill in the art] would understand that it would have at least as many transistors and take up at least as much space as a conventional full-precision arithmetic unit, even when operating in a reduced-precision mode. Indeed, because Dockser's execution unit includes additional control circuits for selecting reduced-precision modes, a [person of ordinary skill in the art] would expect it to be *larger* than a conventional full-precision execution unit, making it unsuitable for scaling, and therefore unsuitable for use in parallel processing arrays.

See Ex. 2051 ¶ 51. Petitioner's asserted combination of incorporating Dockser's FPP into each of multiple PEs in MacMillan's device "would require additional circuitry and chip space and would therefore *increase*

71

IPR2021-00179
Patent 8,407,273 B2

costs," "defeat[ing] MacMillan's stated objective of achieving a highly

parallel SIMD computer architecture at 'lower system cost.'" *See id.*

¶¶ 103–104. Again, MacMillan specifically counsels that to achieve its cost

objectives, "SIMD capabilities should not add significant complexity to the

architecture of a computer system for personal use." Ex. 1009, col. 5,

ll. 42–44. The combination proposed by Petitioner would do so.

MacMillan's PEs already perform full precision operations; Dockser's

selectable subprecision circuitry would add complexity beyond that of a

typical full precision execution unit. *See* Ex. 2051 ¶¶ 103–104; Ex. 1009,

col. 12, ll. 47–49 ("Each PE contains a 32-bit wide data path and can

perform atomic operations on . . . 32-bit words.").

A person of ordinary skill in the art, reading MacMillan and

understanding its emphasis on avoiding increased costs associated with

added complexity and size, would not have been motivated to modify

MacMillan's device by incorporating Dockser's FPP into the PEs. *See*

Chemours Co. FC, LLC v. Daikin Indus., Ltd., 4 F.4th 1370, 1376–77

(Fed. Cir. 2021) (finding insufficient motivation to "increase [a prior art

reference's] melt flow rate to the claimed range," in part because the

reference "includes numerous examples of processing techniques that are

typically used to increase melt flow rate, which [the reference] cautions

should *not* be used due to the risk of obtaining a broader molecular weight

distribution"); *TriVascular*, 812 F.3d at 1067–69 (finding insufficient

motivation to combine two references by "substitut[ing] the recessed barbs

of [one reference] with the protuberances of [another reference], since [the

petitioner's] proposed substitution would destroy the basic objective of the

barbs"); *In re Gordon*, 733 F.2d 900, 902 (Fed. Cir. 1984) (finding no reason

to modify a prior art device where the modification would render the device

72

IPR2021-00179
Patent 8,407,273 B2

"inoperable for its intended purpose"); *see also Henny Penny*, 938 F.3d at
1332 (affirming the Board's finding of no motivation to combine where the
Board "credited [the patent owner's] expert's testimony that following [one
reference's] method of diverting and cooling the oil in [another reference's]
system would introduce 'additional plumbing and complexity'"). Notably,
this is not merely an economic concern over how expensive it would be to
incorporate Dockser's FPP into MacMillan's PEs; the evidence shows that
doing so would increase costs because of the added complexity and size of
the FPP—a concern that MacMillan explicitly counsels against. *See*
Sur-Reply 15.

Petitioner's position is that the individual benefits of MacMillan's
approach (increasing performance speed from many low-cost processors
operating in parallel) and Dockser's approach (lowering power consumption
using additional circuitry added to a single, more complex processor) would
have motivated a person of ordinary skill in the art to combine the
references' teachings by incorporating Dockser's FPP into the multiple PEs
in MacMillan's device. Pet. 5, 47–48. Given the express teachings of
MacMillan, however, we are not persuaded that an ordinarily skilled artisan
would have been motivated to combine the two different approaches in that
manner. Specifically, MacMillan's repeated statements that its approach is
designed to avoid increased costs from added complexity and size (and
thereby allow for many low-cost processors in the device) would have
dissuaded a person of ordinary skill in the art from doing so.

To the extent Petitioner and Mr. Goodin fault Patent Owner for not
providing evidence supporting its position, we are not persuaded. *See* Reply
9–10; Ex. 1071 ¶ 31. Petitioner bears the burden of proving unpatentability
by a preponderance of the evidence, which includes proving that a person of

73

IPR2021-00179
Patent 8,407,273 B2

ordinary skill in the art would have had reason to combine the prior art references. *See* 35 U.S.C. § 316(e); *Magnum Oil*, 829 F.3d at 1380. Moreover, Patent Owner's arguments are supported by evidence. Patent Owner relies on the disclosures of Dockser and MacMillan themselves, expert testimony from Dr. Khatri (which we credit on the issue of reasons to combine Dockser and MacMillan), and cross-examination testimony from Mr. Goodin. *See* PO Resp. 34–36; Sur-Reply 14–15; Ex. 2051 ¶¶ 101–104.

For the reasons explained above, we are not persuaded that a person of ordinary skill in the art would have been motivated to combine the teachings of Dockser and MacMillan in the manner asserted by Petitioner.[8] Accordingly, Petitioner has not shown, by a preponderance of the evidence, that claim 8 would have been obvious based on Dockser and MacMillan under 35 U.S.C. § 103(a).

3. *Claims 3–7, 9–20, 25, 36–61, and 63*

Many of the other challenged claims include limitations similar to the "exceeds" limitation of claim 8. Claims 5 and 10 recite the same limitation, replacing "at least one hundred" with "at least ten" and "at least five hundred," respectively. Claim 36 recites that "the number of LPHDR execution units in the device exceeds the non-negative integer number of

---

[8] Given this determination, we need not address Patent Owner's argument that the references fail to teach the "exceeds" limitation of claim 8 (as well as similar limitations in claims 5, 6, 10–18, 36–61, and 63) because the alleged "LPHDR execution units" (i.e., the Dockser FPPs in the PEs) are "adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide" (in addition to executing at reduced precision) and thus "the number of LPHDR execution units will never exceed the number of full-precision multiplication execution units," as required by the claim. *See* PO Resp. 27–34; Reply 12–15; Sur-Reply 10–14.

IPR2021-00179
Patent 8,407,273 B2

execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide." Claims 40, 43, and 45, which depend from claim 36, mirror claims 5, 8, and 10. Claim 6 depends from claim 5, claims 11–18 depend from claim 8, and claims 37–61 and 63 depend from claim 36. For all of these claims, Petitioner relies on the same theory regarding Dockser and MacMillan asserted for claim 8—namely, incorporating Dockser's FPP into each PE of MacMillan operating in parallel, such that the number of "LPHDR execution units" (i.e., the Dockser FPPs in the PEs) in the combined device allegedly would exceed by a certain amount (e.g., at least one, at least ten, at least 100, at least 500) the number of "execution units . . . adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide" (i.e., the Host CPU floating-point unit). Pet. 47–51, 56.

As to the remaining claims 3, 4, 7, 9, 19, 20, and 25 (that do not recite the "exceeds" limitation or a similar limitation), Petitioner's analysis is premised on the same combination of Dockser and MacMillan, with Dockser's FPP incorporated into each PE of MacMillan's SIMD-RAM architecture, such that the combined device has multiple "LPHDR execution units" (i.e., the Dockser FPPs in the PEs) operating in parallel. *Id.* at 47–49, 53–55.

We are not persuaded that a person of ordinary skill in the art would have been motivated to combine the teachings of Dockser and MacMillan in the manner asserted by Petitioner. *See supra* Section II.F.2. Accordingly, Petitioner has not shown, by a preponderance of the evidence, that claims 3–7, 9–20, 25, 36–61, and 63 would have been obvious based on Dockser and MacMillan under 35 U.S.C. § 103(a).

75

IPR2021-00179
Patent 8,407,273 B2

4.   Claims 1, 2, 21–24, 26, and 28

As explained above, we conclude that claims 1, 2, 21–24, 26, and 28 are unpatentable over Dockser, and over Dockser and Tong, under 35 U.S.C. § 103(a). *See supra* Sections II.D–E. As such, we need not address Petitioner's alternative ground based on Dockser and MacMillan as to claims 1, 2, 21–24, 26, and 28. *See Boston Sci. Scimed, Inc. v. Cook Grp. Inc.*, 809 F. App'x 984, 990 (Fed. Cir. Apr. 30, 2020) (non-precedential) (recognizing that "the Board need not address issues that are not necessary to the resolution of the proceeding" and, thus, agreeing that the Board has "discretion to decline to decide additional instituted grounds once the petitioner has prevailed on all its challenged claims").

G.   Obviousness Ground Based on Dockser, Tong, and MacMillan (Claims 1–26, 28, 32–61, 63, and 67–70)

1.   Claim 8

Petitioner contends that claims 1–26, 28, 32–61, 63, and 67–70 are unpatentable over Dockser, Tong, and MacMillan under 35 U.S.C. § 103(a). Pet. 56–61. We again begin with illustrative claim 8. Petitioner presents two theories. First, Petitioner argues that it would have been obvious to operate the FPPs in the Dockser-MacMillan combination at Tong's precision levels, such that "[t]he resulting Dockser/Tong/MacMillan combination (*i.e.*, MacMillan's multi-PE system using Dockser's FPP with as low as 5-fraction-bit precision as taught by Tong)" teaches the claim limitations "for the same reasons Dockser/MacMillan does," referring to Petitioner's earlier analysis of the asserted ground based on Dockser and MacMillan. Pet. 56–57. We are not persuaded by those arguments for the reasons explained above. *See supra* Section II.F.2.

IPR2021-00179
Patent 8,407,273 B2

Second, Petitioner presents another theory based on a "customized implementation" of Dockser's FPPs. Pet. 59–61 (citing Ex. 1003 ¶¶ 418–428); Reply 15–17 (citing Ex. 1071 ¶¶ 56–62). Petitioner argues that a person of ordinary skill in the art "would have been motivated to customize Dockser's FPPs in MacMillan's PEs to only operate at precision levels lower than full FP 32-bit operations, in view of Tong's teachings that 'the fine precision of the 23-bit mantissa is not essential.'" Pet. 59–60 (citing Ex. 1008, 279). Specifically, according to Petitioner, an ordinarily skilled artisan would have been motivated to "implement Dockser's FPPs in the embedded signal-processing system with smaller than 32-bit registers to not waste circuit space or incur unnecessary cost in having some register elements that will always be unpowered" and likewise "implement the multiplier logic in Dockser's FPP to have only as many logic elements as needed to multiply mantissas of the reduced bitwidth (smaller than 23-bit) corresponding to the precision level selected for the embedded application." *Id.* at 60.

Petitioner's analysis is premised on an "alternative interpretation" of the portion of the "exceeds" limitation referring to execution units "adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide." *Id.* at 59–61. Under that interpretation, an execution unit could meet the "adapted to" clause if it is capable of 32-bit multiplication in *some* configurations. *Id.* at 61. Dockser's FPP, modified as described above, would not be such a unit because it has registers and multiplier logic with less than 32 bits, and thus is not capable of 32-bit multiplication in the proposed combination—rather, only the host CPU floating-point unit would meet the "adapted to" clause. *Id.*

IPR2021-00179
Patent 8,407,273 B2

Patent Owner responds that a person of ordinary skill in the art "would not have been motivated to re-engineer the FPP of Dockser to remove its 32-bit capability [based on Tong] and then incorporate the re-engineered Dockser unit into the system of MacMillan." PO Resp. 38–43 (citing Ex. 2051 ¶¶ 108–115). Patent Owner argues that none of the three references support Petitioner's view. *Id.* at 40–43. First, Dockser teaches a device that "support[s] a range of selectable precisions including full precision," and "[a]djusting Dockser by removing its full-precision capability violates a central tenet of Dockser—to always be able to execute full-precision operations." *Id.* at 41–42. Second, Tong similarly teaches a device with "both full- and reduced-precision capabilities," and would have reinforced Dockser's teaching that "while reduced-precision might be a viable option in certain circumstances, an execution unit should retain the ability to operate at full-precision because many applications require full precision." *Id.* at 40–41 (emphasis omitted). Third, the asserted combination, "which requires special, customized registers, logic elements, [and] arithmetic units, . . . would increase manufacturing costs and goes directly against the teachings of MacMillan," in particular its teaching that adding "significant complexity" to the SIMD architecture should be avoided. *Id.* at 42–43. The parties in their Reply and Sur-Reply disagree as to the meaning of certain cited disclosures in the references. *See* Reply 15–17; Sur-Reply 15–18.

After reviewing the parties' arguments and evidence, and the full disclosures of Dockser, Tong, and MacMillan, we agree with Patent Owner. Petitioner's asserted combination is predicated on *removing* what is a fundamental feature of Dockser and Tong—namely, the ability to operate at full precision or select a precision less than full precision. Dockser teaches a

78

IPR2021-00179
Patent 8,407,273 B2

single processor with a default full precision mode that is capable of also operating in selectable reduced precision modes. Ex. 1007 ¶¶ 3–4 ("[T]he common situation is that for certain applications, . . . a reduced precision may be acceptable, and for other applications, . . . a greater precision may be needed. Accordingly, there is a need in the art for a floating-point processor in which the reduced precision, or subprecision, of the floating-point format is selectable."). Dockser's FPP operates on 32-bit full precision values and performs arithmetic operations at a particular subprecision if the associated "subprecision select bits" are set, but if not, operates at full precision. *Id.* ¶¶ 17–19, 23–29, 32, claims 1 (reciting "a floating-point processor having a maximum precision" and "selecting a subprecision less than the maximum precision"), 8, 15, 20. Dockser thus requires full precision with the capability to select a subprecision less than full precision, as Mr. Goodin agrees. *See* Ex. 2043, 36:16–37:17, 38:3–8 ("I don't believe Dockser discloses any embodiments without floating-point . . . selectable precision."), 55:21–57:12 ("Dockser requires selectable precision"). Indeed, Petitioner explicitly relies on the fact that "Dockser's precision is selectable" when asserting that a person of ordinary skill in the art would have combined the teachings of Dockser and Tong. *See* Pet. 41; Reply 8; Sur-Reply 17. We see no basis to conclude that a person of ordinary skill in the art would have been motivated to remove the feature that Dockser describes as always present and that is the basis for Petitioner's other assertions regarding Dockser and Tong. *See Chemours*, 4 F.4th at 1376–77 (finding insufficient motivation to modify a prior art reference "when doing so would necessarily involve altering the inventive concept of" the reference); Public Tr. 41:21–42:8.

IPR2021-00179
Patent 8,407,273 B2

Tong likewise does not support Petitioner's asserted motivation to combine. Tong discloses that "not all programs need the precision provided by generic FP hardware," and "the fine precision of the 23-bit mantissa is not essential" and "a single custom FP format may be a viable option" for particular applications, but full precision is necessary for other applications. *See* Ex. 1008, 278–280. Tong states that "scientific programs such as large-scale computational fluid dynamics or electric circuit simulation . . . require a huge amount of precision." *Id.* at 279. Tong further discloses that

> by reducing either precision or range from 32 to fewer bits, we should be able to create custom FP hardware which has lower power simply because of the bit reductions. For a narrow, application-specific task, a single custom FP format may be a viable option. However, *to be more generally useful, we need to consider arithmetic architectures which can scale to different FP formats*. Even though we may be able to assume that most of our operands can be computed successfully in limited precision, it appears *inevitable that some fraction of our operands will require full IEEE-standard precision*.

Id. at 280 (emphasis added). Given those disclosures, we are not persuaded that a person of ordinary skill in the art would have been motivated to remove from Dockser's FPP the capability (i.e., selectability from full precision) that Tong characterizes as important.[9]

---

[9] Petitioner has shown sufficiently that a person of ordinary skill in the art would have been motivated to *select* a particular precision level in Dockser's FPP based on the teachings of Tong, but not to *remove* functionality from the FPP itself. *See supra* Section II.E.2; *compare* Pet. 41–42 (arguing that Tong "would have motivated a [person of ordinary skill in the art] to *configure* Dockser's FPP . . . to *operate at* a *selected* precision level retaining as few as 5 mantissa fraction bits . . . or others empirically determined to not require greater precision using Tong's techniques" (emphasis added)), *with id.* at 59–60 (arguing that a person of ordinary skill in the art "would have been motivated to *customize* Dockser's FPPs in

80

IPR2021-00179
Patent 8,407,273 B2

Finally, with respect to MacMillan, the reference states that to achieve its cost objectives, "SIMD capabilities should not add significant complexity to the architecture of a computer system for personal use." Ex. 1009, col. 5, ll. 42–44. The Dockser-Tong-MacMillan combined device would add complexity because it involves a "customized implementation" of the multiplier logic in Dockser's FPP. *See* Pet. 59–60; Reply 15. Further, to the extent some selectability remains (among levels below full precision) in Petitioner's "customized implementation," the modified Dockser FPP would still require Dockser's additional control circuitry to implement that selectable subprecision. As explained above, such circuitry would increase costs due to added complexity and size, which MacMillan expressly counsels against. *See supra* Section II.F.2.

We credit the testimony of Dr. Khatri on these points, as it is explained sufficiently and is consistent with the disclosures of the references. *See* Ex. 2051 ¶¶ 46–48, 107–115. Petitioner has not shown sufficiently that a person of ordinary skill in the art would have combined the teachings of Dockser, Tong, and MacMillan to remove from Dockser's FPP the ability to select a precision less than full precision and then incorporate the modified FPP into the PEs of MacMillan. Accordingly, Petitioner has not shown, by a preponderance of the evidence, that claim 8 would have been obvious based on Dockser, Tong, and MacMillan under 35 U.S.C. § 103(a).

MacMillan's PEs" by removing register storage elements to have "smaller than 32-bit registers" (for the register bit-dropping technique) or removing multiplier logic elements (for the multiplier logic technique) (emphasis added)).

IPR2021-00179
Patent 8,407,273 B2

> ### 2. *Claims 3–7, 9–20, 25, 34–61, 63, and 67–70*

Claims 5, 6, 10–18, 35–61, 63, and 67–70 recite (either on their own or via dependency) the "exceeds" limitation or an analogous limitation. *See supra* Section II.F.3. Petitioner relies on the combined teachings of Dockser and MacMillan, advancing its original theory regarding the combination of Dockser and MacMillan as well as the "customized implementation" argument addressed above for certain claims. *See* Pet. 47–61; Reply 15–17. Claims 3, 4, 7, 9, 19, 20, and 25 do not recite the "exceeds" limitation or an analogous limitation, but Petitioner relies on the same arguments regarding the combination of Dockser and MacMillan. *See* Pet. 47–58. Regarding claim 34, Petitioner refers to its earlier analysis of Dockser and MacMillan for claim 3. *Id.* at 58–59. Petitioner does not argue that Tong cures any of the deficiencies identified above regarding the asserted combination of Dockser and MacMillan. *See id.* at 56–61.

We are not persuaded that a person of ordinary skill in the art would have been motivated to combine the teachings of Dockser, Tong, and MacMillan in the manner asserted by Petitioner. *See supra* Sections II.F.2, II.G.1. Accordingly, Petitioner has not shown, by a preponderance of the evidence, that claims 3–7, 9–20, 25, 34–61, 63, and 67–70 would have been obvious based on Dockser, Tong, and MacMillan under 35 U.S.C. § 103(a).

> ### 3. *Claims 1, 2, 21–24, 26, 28, 32, and 33*

As explained above, we conclude that claims 1, 2, 21–24, 26, and 28 are unpatentable over Dockser, and claims 1, 2, 21–24, 26, 28, 32, and 33 are unpatentable over Dockser and Tong, under 35 U.S.C. § 103(a). *See supra* Sections II.D–E. As such, we need not address Petitioner's alternative

IPR2021-00179
Patent 8,407,273 B2

ground based on Dockser, Tong, and MacMillan as to claims 1, 2, 21–24, 26, 28, 32, and 33. *See Boston Sci. Scimed*, 809 F. App'x at 990.

H.   Patent Owner's Motion to Exclude

Patent Owner moves to exclude (1) Exhibits 1073, 1075–1078, 1082–1084, 1088–1090, 1096, and 1097 as "irrelevant and prejudicial" under Federal Rules of Evidence 402 and 403; (2) Exhibit 1082 for the same reasons and also as unauthenticated under Federal Rule of Evidence 901; (3) and Exhibits 1094 and 1095 as containing inadmissible hearsay under Federal Rule of Evidence 802. Mot. 3–7. Patent Owner's Motion is dismissed as moot, as we do not rely on the exhibits in a manner adverse to Patent Owner in our Decision.

In particular, Petitioner cites Exhibits 1073 and 1075 in support of its argument that Dockser and MacMillan teach the "exceeds" limitation of claim 8 and similar limitations in other claims. *See* Mot. 3–4; Reply 14. Petitioner cites Exhibits 1076–1078 and 1097 in response to an argument made by Patent Owner in its Response regarding MacMillan. *See* Mot. 4; Reply 16–17. We need not reach those arguments because we determine that Petitioner has not shown unpatentability of the claims challenged based on the combinations of Dockser and MacMillan, and Dockser, Tong, and MacMillan, for other reasons. *See supra* Sections II.F, II.G.

Petitioner cites Exhibits 1078, 1082–1084, 1088–1090, and 1094–1096 in support of its arguments regarding objective indicia of nonobviousness. *See* Mot. 4–7; Reply 21–22, 25, 27–28. As explained above, we do not find Patent Owner's evidence of objective indicia of nonobviousness persuasive with respect to claims 1, 2, 21–24, 26, 28, 32,

IPR2021-00179
Patent 8,407,273 B2

and 33, and do not rely on the disputed exhibits in reaching that conclusion. *See supra* Sections II.D, II.E.

I.   *Motions to Seal*

The parties move to seal portions of their demonstrative exhibits that quote or describe material that we previously ordered sealed. Papers 52, 54; *see* Papers 27, 48 (granting motions to seal). The parties filed unredacted confidential and redacted public versions of their demonstrative exhibits. Exs. 1098, 2054. Upon reviewing the materials sought to be sealed, it appears that the parties' characterization of the materials is accurate and the redactions are narrowly tailored to only confidential information. The parties have established good cause to seal the redacted portions of the demonstrative exhibits. We grant the Motions to Seal.

III. CONCLUSION[10]

Petitioner has demonstrated, by a preponderance of the evidence, that claims 1, 2, 21–24, 26, 28, 32, and 33 of the '273 patent are unpatentable, but has not demonstrated, by a preponderance of the evidence, that claims 3–20, 25, 34–61, 63, and 67–70 are unpatentable.

---

[10] Should Patent Owner wish to pursue amendment of the challenged claims in a reissue or reexamination proceeding subsequent to the issuance of this Decision, we draw Patent Owner's attention to the April 2019 *Notice Regarding Options for Amendments by Patent Owner Through Reissue or Reexamination During a Pending AIA Trial Proceeding. See* 84 Fed. Reg. 16,654 (Apr. 22, 2019). If Patent Owner chooses to file a reissue application or a request for reexamination of the challenged patent, we remind Patent Owner of its continuing obligation to notify the Board of any such related matters in updated mandatory notices. *See* 37 C.F.R. §§ 42.8(a)(3), 42.8(b)(2).

IPR2021-00179
Patent 8,407,273 B2

In summary:

Claims	35 U.S.C. §	References/ Basis	Claims Shown Unpatentable	Claims Not Shown Unpatentable
1, 2, 21–24, 26, 28	103(a)	Dockser	1, 2, 21–24, 26, 28	
1, 2, 21–24, 26, 28, 32, 33	103(a)	Dockser, Tong	1, 2, 21–24, 26, 28, 32, 33	
1–26, 28, 36–61, 63	103(a)	Dockser, MacMillan[11]		3–20, 25, 36–61, 63
1–26, 28, 32–61, 63, 67–70	103(a)	Dockser, Tong, MacMillan[12]		3–20, 25, 34–61, 63, 67–70
Overall Outcome			1, 2, 21–24, 26, 28, 32, 33	3–20, 25, 34–61, 63, 67–70

IV. ORDER

In consideration of the foregoing, it is hereby:

ORDERED that claims 1, 2, 21–24, 26, 28, 32, and 33 of the ’273 patent have been shown to be unpatentable, and claims 3–20, 25, 34–61, 63, and 67–70 of the ’273 patent have not been shown to be unpatentable;

---

[11] As explained above, given our disposition of the grounds based on Dockser and the combination of Dockser and Tong, we do not reach Petitioner’s alternative ground asserting that claims 1, 2, 21–24, 26, and 28 are unpatentable over Dockser and MacMillan. *See supra* Section II.F.4.

[12] As explained above, given our disposition of the grounds based on Dockser and the combination of Dockser and Tong, we do not reach Petitioner’s alternative ground asserting that claims 1, 2, 21–24, 26, 28, 32, and 33 are unpatentable over Dockser, Tong, and MacMillan. *See supra* Section II.G.3.

IPR2021-00179
Patent 8,407,273 B2

FURTHER ORDERED that Patent Owner's Motion to Exclude (Paper 46) is *dismissed*; and

FURTHER ORDERED that the parties' Motions to Seal (Papers 52 and 54) are *granted*, and the unredacted confidential versions of the demonstrative exhibits (Exhibits 1098 and 2054) shall remain under seal pursuant to the default protective order previously entered in the instant proceeding.

This is a final decision. Parties to the proceeding seeking judicial review of the decision must comply with the notice and service requirements of 37 C.F.R. § 90.2.

IPR2021-00179
Patent 8,407,273 B2

FOR PETITIONER:

Elisabeth H. Hunt
Richard F. Giunta
Anant K. Saraswat
Adam R. Wichman
Nathan R. Speed
WOLF, GREENFIELD & SACKS, P.C.
ehunt-ptab@wolfgreenfield.com
rgiunta-ptab@wolfgreenfield.com
asaraswat-ptab@wolfgreenfield.com

FOR PATENT OWNER:

Peter Lambrianakos
Vincent J. Rubino, III
Enrique W. Iturralde
Richard Cowell
FABRICANT LLP
plambrianakos@fabricantllp.com
vrubino@fabricantllp.com
eiturralde@fabricantllp.com
rcowell@fabricantllp.com

Brian M. Seeve
Matthew D. Vella
PRINCE LOBEL TYE LLP
bseeve@pricelobel.com
mvella@pricelobel.com